

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 14:57:29 ; Search time 21 Seconds
(without alignments)
1617.887 Million cell updates/sec

Title: US-10-034-849-2

Perfect score: 4188

Sequence: 1 MTEVFTVLDSSYEVVKGEP.....LKASATGQXTLFDLAKSK 803

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/prodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/prodata/1/iaa/5A COMB.pep.*
4: /cgn2_6/prodata/1/iaa/5B COMB.pep.*
5: /cgn2_6/prodata/1/iaa/PCRUS COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4188	100.0	803	2	US-08-907-166-4
2	4188	100.0	803	4	US-09-391-340-4
3	3028.5	72.3	803	1	US-08-062-368-4
4	2963.5	70.8	803	1	US-08-062-368-2
5	1563	37.3	788	2	US-08-907-166-6
6	1563	37.3	788	4	US-09-391-340-6
7	1237.5	29.5	774	3	US-08-902-632-2
8	1237.5	29.5	776	2	US-08-688-649-37
9	1236.5	29.5	775	1	US-07-966-278-1
10	1236.5	29.5	775	1	US-08-424-921-1
11	1236.5	29.5	775	2	US-08-556-355A-1
12	1236.5	29.5	775	2	US-07-803-627A-1
13	1236.5	29.5	775	4	US-09-244-889A-1
14	1234.5	29.5	774	3	US-09-073-354-1
15	1234.5	29.5	774	3	US-08-656-005A-1
16	1234.5	29.5	774	3	US-09-073-259-1
17	1234.5	29.5	774	3	US-09-363-095-1
18	1234.5	29.5	774	3	US-09-418-027-1
19	1216.5	29.0	773	4	US-09-585-858-35
20	1206.5	28.8	727	4	US-09-585-858-32
21	1146.5	27.4	778	2	US-08-906-925-4
22	1104	26.4	779	1	US-08-375-134-12
23	1104	26.4	779	5	PCT-US95-15263-12
24	760.5	18.2	1107	1	US-08-366-577-2
25	760.5	18.2	1107	5	PCT-US96-00005-2
26	739.5	17.7	1019	1	US-08-271-364A-7
27	739.5	17.7	1019	2	US-08-222-715B-26

28	722.5	17.3	734	4	US-09-585-858-8	Sequence 8, Appli
29	696.5	16.6	738	4	US-09-585-858-6	Sequence 6, Appli
30	693.5	16.6	738	4	US-09-585-858-7	Sequence 7, Appli
31	661.5	15.8	741	4	US-09-585-858-11	Sequence 11, Appli
32	654	15.6	735	4	US-09-585-858-10	Sequence 10, Appli
33	648.5	15.5	734	4	US-09-585-858-9	Sequence 9, Appli
34	635	15.2	1022	1	US-08-271-364A-8	Sequence 8, Appli
35	635	15.2	1022	2	US-08-222-715B-27	Sequence 27, Appli
36	579.5	13.8	757	4	US-09-585-858-23	Sequence 23, Appli
37	579.5	13.8	1462	3	US-07-792-600-31	Sequence 31, Appli
38	579.5	13.8	1462	3	US-09-157-021-31	Sequence 31, Appli
39	579.5	13.8	1462	3	US-09-156-842-31	Sequence 31, Appli
40	564.5	13.5	757	4	US-09-585-858-24	Sequence 24, Appli
41	560.5	13.4	702	4	US-09-585-858-33	Sequence 33, Appli
42	542.5	13.0	877	2	US-08-907-166-8	Sequence 8, Appli
43	542.5	13.0	877	4	US-09-391-340-8	Sequence 8, Appli
44	526.5	12.6	1026	4	US-09-252-991A-30231	Sequence 30231, A
45	524.5	12.5	1015	2	US-08-680-326-32	Sequence 32, Appli

ALIGNMENTS

RESULT 1

US-08-907-166-4

; Sequence 4, Application US/08907166

; Patent No. 5948666

; GENERAL INFORMATION:

; APPLICANT: Callen, Walter

; APPLICANT: Mather, Eric

; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES

; FILE REFERENCE: 09010/027001

; CURRENT APPLICATION NUMBER: US/08/907,166

; CURRENT FILING DATE: 1997-08-06

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 4

; LENGTH: 803

; TYPE: PRT

; ORGANISM: Pyrolobus fumarius

US-08-907-166-4

Query Match	100.0%	Score 4188;	DB 2;	Length 803;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 803;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MTEVFTVLDSSYEVVKGEPQVIINGIAENGERSVVLIDRSFRPYFALLAAGADPKQVAQ	60	
DB	1	MTEVFTVLDSSYEVVKGEPQVIINGIAENGERSVVLIDRSFRPYFALLAAGADPKQVAQ	60	
QY	61	RIRALSREKSPITIGVEDDKRKYFGPRVLRITVLPVAVREYRELKQVGVGVLEAD	120	
DB	61	RIRALSREKSPITIGVEDDKRKYFGPRVLRITVLPVAVREYRELKQVGVGVLEAD	120	
QY	121	IRFAMRYLIDHDLFFPTWYRVEAEPLKMGFRVDKVLVKSRLPEPLAGEALPTKL	180	
DB	121	IRFAMRYLIDHDLFFPTWYRVEAEPLKMGFRVDKVLVKSRLPEPLAGEALPTKL	180	
QY	181	RILAPDIIVYKQSGSPRPERDPIVIAVKTDGDEVLFIAEGKDDRKPIREFVEYVKRYD	240	
DB	181	RILAPDIIVYKQSGSPRPERDPIVIAVKTDGDEVLFIAEGKDDRKPIREFVEYVKRYD	240	
QY	241	PDIIIVGYNHNFDPYLLRRARILGKLDVTRRGABETTSVHGHSVFGSLNVDLYA	300	
DB	241	PDIIIVGYNHNFDPYLLRRARILGKLDVTRRGABETTSVHGHSVFGSLNVDLYA	300	
QY	301	REMPRIKISLEEAVALYGVNKKSERVLINWEPDYWDKPKRPLALQYARDVRYTGY	360	
DB	301	REMPRIKISLEEAVALYGVNKKSERVLINWEPDYWDKPKRPLALQYARDVRYTGY	360	
QY	361	LAEKILPFAIQLSYVTGLPLDQVGAMSVGFRLWYLRFAAFKMKELVNRVERPEYVYG	420	
DB	361	LAEKILPFAIQLSYVTGLPLDQVGAMSVGFRLWYLRFAAFKMKELVNRVERPEYVYG	420	

Qy	421	AIVLEPLRGVHENIAVLDPSSMYPMIMIKYNGPDTLVRPCCKGCGCWAEAPVKHRR	480
Db	421	AIVLEPLRGVHENIAVLDPSSMYPMIMIKYNGPDTLVRPCCKGCGCWAEAPVKHRR	480
Qy	481	RCPGGFKTVLERLELRKRVRAEMKKYPDPSPEYRLDLDERQALKVLANASYGYMGWS	540
Db	481	RCPGGFKTVLERLELRKRVRAEMKKYPDPSPEYRLDLDERQALKVLANASYGYMGWS	540
Qy	541	ARMYCCEKAVTANGRHLLIRTAINIAERKLGVLYGDTDSLFTVYDPEKVENFIKIIE	600
Db	541	ARMYCCEKAVTANGRHLLIRTAINIAERKLGVLYGDTDSLFTVYDPEKVENFIKIIE	600
Qy	601	ELGFEEKLVKYRILFFTEAKRYAGLLEDGRIDIVGFEAVRGDWCELAKEVOTKVEIV	660
Db	601	ELGFEEKLVKYRILFFTEAKRYAGLLEDGRIDIVGFEAVRGDWCELAKEVOTKVEIV	660
Qy	661	LKTSEYNKAVEYRKIVVELBEGKVPTEKLVIWTKLSKRLEEYTTBAPHVAAKRMLSAG	720
Db	661	LKTSEYNKAVEYRKIVVELBEGKVPTEKLVIWTKLSKRLEEYTTBAPHVAAKRMLSAG	720
Qy	721	YRVSPGDKIGYIVKGGRISQRAMPYPMWKDPSQIDVTYYVDHOIIPALRLTLGVFGIT	780
Db	721	YRVSPGDKIGYIVKGGRISQRAMPYPMWKDPSQIDVTYYVDHOIIPALRLTLGVFGIT	780
Qy	781	EKKLKASATGQKTLDFFLAKSKS	803
Db	781	EKKLKASATGQKTLDFFLAKSKS	803

RESULT 2

US-09-391-340-4
; Sequence 4, Application US/09391340A
; Patent No. 6492511

```

; GENERAL INFORMATION:
; APPLICANT: Callen, Walter
; APPLICANT: Macher, Eric
; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
; FILE REFERENCE: 09010/027001
; CURRENT APPLICATION NUMBER: US/09/391,340A
; CURRENT FILING DATE: 1999-09-07
; EARLIER APPLICATION NUMBER: US 08/907,166
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 803
; TYPE: PRT
; ORGANISM: Pyrolobus fumarius
US-09-391-340-4

```

Query Match	100.0%;	Score 4188;	DB 4;	Length 803;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 803;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MTEVFTVLDDSSYEYVKGKEQVLIINGTGAENGSRVWLIDRSFRPYFYALLAPGADPKQVAAQ	60	
Db	1	MTEVFTVLDDSSYEYVKGKEQVLIINGTGAENGSRVWLIDRSFRPYFYALLAPGADPKQVAAQ	60	
QY	61	RIRALSRRPKSPIIGVEDDKKFKYGRPRRVLRIITVLPEAVREYRELKVNVDGVEDVLEAD	120	
Db	61	RIRALSRRPKSPIIGVEDDKKFKYGRPRRVLRIITVLPEAVREYRELKVNVDGVEDVLEAD	120	
QY	121	IRFAMRYLIIDHDLFFFTWYVEAEPLKNGKGFVDKVLVKSRPPELYGEALATPKLPDL	180	
Db	121	IRFAMRYLIIDHDLFFFTWYVEAEPLKNGKGFVDKVLVKSRPPELYGEALATPKLPDL	180	
QY	181	RILAFDIEVYSKQSGSRPRDPPIVIAKTTDGDDEVLFIAEGDKDKRIRSEFVYVKRYD	240	
Db	181	RILAFDIEVYSKQSGSRPRDPPIVIAKTTDGDDEVLFIAEGDKDKRIRSEFVYVKRYD	240	
QY	241	PDIIIVGYNHEDWPPYLIIRARILGTLDVTRRVGAEPPTSIVGHVSVYGBRLNVLDLYDA	300	
:				

241	Db	PDIIVGNNHDPVILRRARILGIKUDVTRVGAEBETTSVGHSHSVPGRLNVLDYA	300
301	Qy	EMPEIKIKISLEBAEYILGVMMKXSERVIINWWEIPDYWDPPKKRPLLQVARDVRA	360
301	Db	EMPEIKIKISLEBAEYILGVMMKXSERVIINWWEIPDYWDPPKKRPLLQVARDVRA	360
361	Qy	LAEKILPFAIOISYVTGLPLDQVGAMSVGFLEWYLIRAAFKMKELVPNRPES	420
361	Db	LAEKILPFAIOISYVTGLPLDQVGAMSVGFLEWYLIRAAFKMKELVPNRPES	420
421	Qy	AIVLEPLRGVHENIAVLDFSSWYNIMIKYNNVGPDTLVREGEKCGEGC	480
421	Db	AIVLEPLRGVHENIAVLDFSSWYNIMIKYNNVGPDTLVREGEKCGEGC	480
481	Qy	RCPPGFFKTVLERLLELRKRVRAEMKKYPPDPSPEYRLDLDERQALKVLANAS	540
481	Db	RCPPGFFKTVLERLLELRKRVRAEMKKYPPDPSPEYRLDLDERQALKVLANAS	540
541	Qy	ARWYCRECAKAVTANGRHILRTAINIARKLGKVIYIGDTSLEVTYDPEKVENFI	600
541	Db	ARWYCRECAKAVTANGRHILRTAINIARKLGKVIYIGDTSLEVTYDPEKVENFI	600
601	Qy	ELGFEIKLEKVKYKRLFFTEAKRVYAGLLEBGRIDIVGFGEAVRGDMWCELAKE	660
601	Db	ELGFEIKLEKVKYKRLFFTEAKRVYAGLLEBGRIDIVGFGEAVRGDMWCELAKE	660
661	Qy	LKTSSEVNVKAVEYVRKI VKELEEGKVPTEKLVINKTLSKRLLEEYTTT	720
661	Db	LKTSSEVNVKAVEYVRKI VKELEEGKVPTEKLVINKTLSKRLLEEYTTT	720
721	Qy	YRVSPGDKIGYIVVKGGGRIISORAWPYFMWKDPSQIDVTYVVDHQIIP	780
721	Db	YRVSPGDKIGYIVVKGGGRIISORAWPYFMWKDPSQIDVTYVVDHQIIP	780
781	Qy	EKKLKASATGOKTFLDFLAKKSK 803	
781	Db	EKKLKASATGOKTFLDFLAKKSK 803	

RESULT 3

US-08-062-368-4
 US-Sequence 4, Application US/08062368
 Patent No. 5491086
 GENERAL INFORMATION:
 APPLICANT: Gelfand, David H.
 APPLICANT: Wang, Alice M.
 TITLE OF INVENTION: Purified Thermostable Nucleic Acid
 TITLE OF INVENTION: Polymerases Enzyme From Pyrodictum Species
 NUMBER OF SEQUENCES: 35
 CORRESPONDENCE ADDRESS:
 ADDRESS: Hoffmann-La Roche Inc.
 STREET: 340 Kingsland Street
 CITY: Nutley
 STATE: New Jersey
 COUNTRY: U.S.A.
 ZIP: 07110
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/062,368
 FILING DATE: 19930514
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Sias Ph.D., Stacey R.
 REGISTRATION NUMBER: 32,630
 REFERENCE/DOCKET NUMBER: 8584
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (510) 814-2863
 TELEFAX: (510) 814-2977
 INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 803 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-062-368-4

Query Match 72.3%; Score 3028.5; DB 1; Length 803;
Best Local Similarity 70.8%; Pred. No. 2.3e-270;
Matches 570; Conservative 106; Mismatches 122; Indels 7; Gaps 5;

QY 1 MTEVV-FTVLDSSYEVGKEPOVILWGIAENGERVVLIDRSFRPYFALLAPADP--KQ 57
DB 1 MTEIEFVLDSYELLGKEPVILMGITLDGKRVLLDHRFPYFALYARGYEDVVE 60
QY 58 VAQIRALSRKPSPIIGVEDDKRYFGRPRVLRIRTVLPEAVREYRELKVNVDGVDEL 117
DB 61 IASIRLSVVKSPIIDAKPLDKRYFGRPKAVKITTWIPESVRHYREAVKKEGVEDSL 120
QY 118 EADIRFAMRYLIDHDLFPFTWYRVEAEPLNKNMGRFVKVYLKSRPEPLYGBALAPTKL 177
DB 121 EADIRFAMRYLIDKRLPYFTVYRIPVEDAGRNFGFRVDRYKVAGDPPEL--ADITRIDL 178
QY 178 PDRLILAFDIYVSKQSPRPRDPTVIAVKTDDGDEVLFIAEGKDDKRPREFVEYVK 237
DB 179 PMRLVAFDIYVSRGSPNRPDPVILVSRDSBEGKERLIEAEGHDDRRLREFVEYVR 238
QY 238 RYDPDIIVGNNHFDWPIILRRARILIGIKLOVTRRVGAEPSTTVSHGVSVQRLNVLDY 297
DB 239 AFDPDIIVGNSHFDWPIILMERARRLGINLDTTRRVGAEPSTTVSHGVSVQRLNVLDY 298
QY 298 DYAEEMPEIKSLLEEVAEYLGVMKKSERVIINWMEIPDYMDPKRPLLLQYARDVRA 357
DB 299 DYAEEMPEIKMKTLEEVAEYLGVMKKSERVIINWMEIPDYMDDEKKRQLERVALDVRA 358
QY 358 TYGLAEKILPFAIQLSVTGVLDDQVGMVSGFRLEWYLRRAAPKMKELVPRVERPEET 417
DB 359 TYGLAEKMLPFAIQLSVTGVLDDQVGMVSGFRLEWYLRRAAYDNKELVPRVERRGES 418
QY 418 YRGAIVLPLRGVHENTAVLDFSSMYENIMIKYNVGPDTLVRPGEKGEC--GWEAPEVK 476
DB 419 YKGAIVLPLKGVHENVVVLDFSSMYENIMIKYNVGPDTLVDDPSECPKYGCVYAPEVG 478
QY 477 HRFRRCPPGFKTVLEKLELRKVRBAEMKKYPPDSPEYRLDLDERQKALKVLANASYGM 536
DB 479 HRFRSPGPFKTVLENLLKLRQVKEMKEFPDPSPEYRLYDERQKALKVLANASYGM 538
QY 537 GWSGARWYCRECAKAVTANGPHLIRTAINTARKLGLKVIYGDGDSLVYDPEKVENFIK 596
DB 539 GWSHARWYKRCBAEAVTANGRNILTLTAIEYARKLGLKVIYGDGDSLVYDPEKVEKLE 598
QY 597 IIEELGFEIKLEKVKYKRLFFTEAKRYAGLLEDGRIDIVGFAVRGDMCELAKYQTKV 656
DB 599 FVEKELGFEIKIDIKYKVFTEAKRYAGLLEDGRIDIVGFAVRGDMCELAKYQBEKA 658
QY 657 VEIVLKTSEYKAVYKRVKIVKLEBCKVPIKLVITWKTLSKLEEYTTAEAPHVAARM 716
DB 659 AEIVLNTGNVDKAKISYIREVIKQLREGKVPITKLIITWKTLSKRIEYEHDAHVMAARM 718
QY 717 LSAGYRVSPOGKIGYVIVKGGGRISQRAWPYFVWQDPSQIDVYVYDHOHIIIPAALILGY 776
DB 719 KEAGYEVSPDGKGVYIVKSGSVSSRAYPFYFMV--DPSTIDVNYIYDHOHIVRAALRILSY 777
QY 777 FGITEKVLKASATQKTLDFDLAKK 801
DB 778 FGITEKVLKAAATVQSRCLDFDLAKK 802

RESULT 4
US-08-062-368-2
; Sequence 2, Application US/08062368
; Patent No. 5491086

GENERAL INFORMATION:
APPLICANT: Gelfand, David H.
APPLICANT: Wang, Alice M.
TITLE OF INVENTION: Purified Thermostable Nucleic Acid
TITLE OF INVENTION: Polymerases Enzyme From Pyrodicticum Species
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/062,368
FILING DATE: 19930514
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sias Ph.D., Stacey R.
REGISTRATION NUMBER: 32,630
REFERENCE/DOCKET NUMBER: 8584
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2863
TELEFAX: (510) 814-2977
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 803 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-062-368-2

Query Match 70.8%; Score 2963.5; DB 1; Length 803;
Best Local Similarity 68.4%; Pred. No. 2.3e-264;
Matches 551; Conservative 118; Mismatches 129; Indels 7; Gaps 5;

QY 1 MTEVV-FTVLDSSYEVGKEPOVILWGIAENGERVVLIDRSFRPYFALLAPADP--KQ 57
DB 1 MPEIEFVLDSYELLGKEPVILMGITLDGKRVLLDHRFPYFALYARGYEDVVE 60
QY 58 VAQIRALSRKPSPIIGVEDDKRYFGRPRVLRIRTVLPEAVREYRELKVNVDGVDEL 117
DB 61 VVAARLSMAKSPILIEAKVSKYFGRPKAVKITTWIPESVRHYREAVKKEGVEDSL 120
QY 118 EADIRFAMRYLIDHDLFPFTWYRVEAEPLNKNMGRFVKVYLKSRPEPLYGEALAPTKL 177
DB 121 EADIRFAMRYLIDKRLPYFTVYRVEAEPLNKNMGRFVKVYLKSRPEPLYGEALAPTKL 178
QY 178 PDRLILAFDIYVSKQSPRPRDPTVIAVKTDDGDEVLFIAEGKDDKRPREFVEYVK 237
DB 179 PMRLVAFDIYVSRGSPNRPDPVILVSRDSBEGKERLIEAEGHDDRRLREFVEYVR 238
QY 238 RYDPDIIVGNNHFDWPIILRRARILIGIKLOVTRRVGAEPSTTVSHGVSVQRLNVLDY 297
DB 239 AFDPDIIVGNSHFDWPIILMERARRLGINLDTTRRVGAEPSTTVSHGVSVQRLNVLDY 298
QY 298 DYAEEMPEIKSLLEEVAEYLGVMKKSERVIINWMEIPDYMDPKRPLLLQYARDVRA 357
DB 299 DYAEEMPEIKMKTLEEVAEYLGVMKKSERVIINWMEIPDYMDDEKKRQLERVALDVRA 358
QY 358 TYGLAEKILPFAIQLSVTGVLDDQVGMVSGFRLEWYLRRAAPKMKELVPRVERPEET 417
DB 359 TYGLAEKILPFAIQLSVTGVLDDQVGMVSGFRLEWYLRRAAHDMELVPRVERRES 418
QY 418 YRGAIVLPLRGVHENTAVLDFSSMYENIMIKYNVGPDTLVRPGEKGEC--GWEAPEVK 476
DB 419 YKGAIVLPLKGVHENVVVLDFSSMYENIMIKYNVGPDTLVDDPSECPKYGCVYAPEVG 478

QY	477	HFRRCRCPGFFXTVLRLLELRKRVRAEMKKYPPDSPYRLLDEROKALVLNANASYGM	536
Db	479	HMRSPSGFFXTVLENIJLRKVQREKWEFFPDSPFYIYDEROKALVLNANASYGM	538
QY	537	GHSGARWYCRCAKAVTAWGRHLIRTAINIARKLGLKVIYGDTSLSFVTVDPPEKVENFIK	596
Db	539	GMVHARWYCKRCAEAVTAWGRNLLLSALEYARKLGLVIYGDTSLSFVTVDEKVKLLIE	598
QY	597	LKEELGFPIKLEKTVKYLFTFEAKRVAGLLEDRIDIVGFBAVRGDMCELAKAVEQVK	656
Db	599	FVEQKGFPIKIDKYVKVFFTEAKRVYGLLEDRMDIVGFBAVRGDMCELAKAVEQVK	658
QY	657	VEIILKTSBWNKAVYVRKIYKELEEGKVPTEKLIWIKTLTKSLREYVTTAEAHVVAAKRM	716
Db	659	ABIIILKTGINRAISYIEVRVKUREGIPITKULVIWIKTLTKRIEYEAHPVTAARM	718
QY	717	LSAGYRVSPGDKIGYVIVKGGGRISQRAWPFYMKDPSQDVTYVVDHQLIIPAALRILCY	776
Db	719	KEAGYDVAPGDKIGYIIVKGHGSISRRAYPFMY-DSSKVDTEYYIDHQIVPAMRILSY	777
QY	777	FGITEKKUKASATGOKTLFDFFLAKK	801
Db	778	FGVTEKQLKAASSGHRSLFDFDFAAK	802

```

RESULT 5
US-08-907-166-6
; Sequence 6, Application US/08907166
; Patent No. 5948666
; GENERAL INFORMATION:
; APPLICANT: Callen, Walter
; APPLICANT: Mather, Eric
; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
; FILE REFERENCE: 090107027001
; CURRENT APPLICATION NUMBER: US/08/907,166
; CURRENT FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 788
; TYPE: PRT
; ORGANISM: Archaeoglobus lithotrophicus
US-08-907-166-6

```

Query Match	37.3%;	Score	1563;	DB 2;	Length	788;			
Best Local Similarity	41.3%;	Pred. No.	3.5e-135;						
Matches	338;	Conservative	154;	Mismatches	273;	Indels	54;	Gaps	18;

QY	1	MTEVFTVLDSSEYWGKEPQVLIWGIAENGERWLIIDRSFRPYFYFALLAPGADPKQVQAQ	60
Db	1	MIKVGWMLLDADYITENDRAVIRLWCKDESG-IFAIYDHSFQYFYFALKESGITRADIIVK	59
QY	61	-RIRALSAPKSPITGVBEDKKRYFGRRPRVLRITVLPBAVREYRELKVNQDVGVDVLEA	119
Db	60	IKVQTKKEVITP-LKVBETTAKNLGREVEVFKIYARHPQHPVKLRVVSQ---YLEITREA	115
QY	120	DIRAMRYLIDHDLFPPTWTRVBAEPLNMGFRVDKVLVK--SRPEPLYGEALAPTKL	177
Db	116	DIPFAYRLIDKNLACMGVVIEGVERREK-GLRCVEIKRIERDSQD-----F	163
QY	178	PDRLILAFDIEVYSKQSGSPRRDPPIVIAKVDDCGDEVLFIAQCKDKRKPIREPEVVK	237
Db	164	PELKVAADFCEMLSEVGMGPPEKDPPIIVISIKSGEYEIL--NGDNRELLTRFVKIIR	220
QY	238	RYDPPDIIVGNYNHFDMPYLLRRARILGIKLDVTRVGAEPITSVHGCHVSPGRLINVDLY	297
Db	221	DIDPDIIVGNQDSFDMPYIKKRAEKLVRKLDITGR--DRSELAIIRGSRPKIAGRLINVDLY	278
QY	298	DYASEMPEIKTLEEVAEYLVGMVKKSERVINNWBEIPDYWDPKPKRPLLQYAADDVRA	357
Db	279	DIAMRSLDVKKLVENVAEELG--KKLELADIEAKDIYKHWTSGDRESVIYKSRQDILH	335

QY	358	TYGLAEKILPFAIQUSVYTGPLDQVGAMSVGRLEWYLIRAAFKMKELVYNNRVERPEET	417
Db	336	TYFIAEELPMPHYELSMRIRIPLDQWTRSGRGQVEMILLSEAHKLGELAPNPREM-ADS	394
QY	418	YRGATVLEPLRGVHENIAVLDPSSWYNIMIKYNGPDTLVPGEKCGECGCWEAPEVKH	477
Db	395	YEGAFVLEPARGLHENVLCLOPASWPSIMSYNISPDPLVI--GRCCDCNV--APEVGH	450
QY	478	RPRCPGPFYKTVLERLELRKRVRAEMKKYPPDSPYRLLDROKALKVLANASYGYMG	537
Db	451	KPRKHPDGFYKRIKQMLTEKREIKVKVMKTLDYNSPEYKLDLQATLKVLTNSFYGYG	510
QY	538	WSGARWYCRECAKAVTANGRHILPAINIARKLGLKVIYGTDSLFVYDPEKVENFIKI	597
Db	511	WSLARWYCKEACATANGRHIFIKTSARIAKLEFVLYGTDSIFVXKQGLSLEELKE	570
QY	598	IK-----BELGEFIEKLEKVVYKRLFFTEAKKRYAGLLEDGRIDWVGFAVRGDMCELAK	650
Db	571	VKKLIGKLSEMPQIEIDEYVETIFFVE-KKSYAGLTQDGRILVKGLEVRGDMCELAK	629
QY	651	EVQTKVVEIIVLKTSEVKNKAVYRKYVKSLEBEGKVPTEKLVITWYLSKRLBETTTAPHV	710
Db	630	KIQKGVTEIILKXNPKAAEYVKGVIIEIKAGKIPLEDYIYIKGLTRKPKSYKESMQAHV	689
QY	711	VAACKMLSAGYRVSPGDKIGVIVKGGGRIISORAMPYFMVKD-----PSQIDV	758
Db	690	KAAMKAARGIVYITGSKVGFVYKGVGNIGRAFPFSDLIIEFDGEVITDLOGNKYKIDK	749
QY	759	TYVDHQIILPAALRIILVFGITEKKLKASATGQKTLFDF	797
Db	750	EYIDHQVLPVLRILRFRFYTEAQLKGAAS-QOITLDAF	787

```

RESULT 6
US-09-391-340-6
; Sequence 6, Application US/09391340A
; Patent No. 6492511
; GENERAL INFORMATION:
; APPLICANT: Callen, Walter
; APPLICANT: Mathen, Eric
; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
; FILE REFERENCE: 03010/027001
; CURRENT APPLICATION NUMBER: US/09/391,340A
; CURRENT FILING DATE: 1999-03-07
; EARLIER APPLICATION NUMBER: US 08/907,166
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 788
; TYPE: PRT
; ORGANISM: Archaeoglobus lithotrophicus
US-09-391-340-6

```

[illegible]

QY 238 RYDDIIVGVNNHFDWPIYLLRRARILGIKLDVTRVGABPTTSVHGHSVGPRLNVLY 297
Db 221 DIDPDIIVGNQDSFDPYIKKRAELKVLKLDIGR--DRSELAIRGPRPKIAGELNVLY 278
QY 298 DYAEEMPEIKIKSLEAEVGVNKKSERVIVNWEIPDYDDPKRPLLLQYARDVRA 357
Db 279 DIAMSLDVKKLENVAEFLG--KKELADIKAQIYKHTS--GRESVIKYSRQDILH 335
QY 358 TYGLAEKILPAIQLSVYVGLPQGVGMSVGRLEWLYIRAFKMKELVFNVERPEET 417
Db 336 TYPIABELLMHVELSMRIPLDVDRSGRGQVWLLSEAHKLGELAPNPREM-ADS 394
QY 418 YRGAIVLEPLRGVHENTAVLDFSSMYNIMIKNVGPDILVRFGKCGCGCWEAEVRKH 477
Db 395 YEGAFVLEPARGHLENVICLDFASMYPSIMISYNSPDTLVI--GKDDCNV--APEVGH 450
QY 478 RFRCPGPFKTVLERLELRKVRAMKKYPPDPSPEYRLDLDERQKALKVLANASYCYMG 537
Db 451 KFRKHPDGFPRILKMLIEKRREIKVMKTLDYNSPEYKLLDIKQATLKVLTSFYGTG 510
QY 538 WSGARWYCRCAKAVTANGRLHIRTAINIKULGLKVIYGDTSLSFVYDPEKVENFIKI 597
Db 511 WSLARWYCKEACATTANGRHFIKTSARIAKELGFEVLYGDTDSIFVKKDGLSLEBKKE 570
QY 598 IK-----BELGPEIKLEKVKYKELFFTEAKKRYAGLLEDGRIDIVGFEAVRGDWCELAK 650
Db 571 VKKLGKLSMPQIHSIDYEYETIFFVE--KKRYAGTQDGRIVVKGLEVRGDCWCELAK 629
QY 651 EVQTKVVEIVLKTSEVNKAVYVRKIVKLEBEGKVPTEKLVINWTKLSKLEEYTTBAPHV 710
Db 630 KIQKGVIEIILKEKNPEKAAEYVKGVIEIKAGKIPLEDYIYKGLTRKESKYESMQAHV 689
QY 711 VAARMKLSAGYRVSPODKIGVIVKGGGRISQRAWPYFMKD-----PSQIDV 758
Db 690 KAAMKAARKGIVTVTIGSKVGVVTKGVGNIGDRAPPSDLIEDFDGEVITOLDGNKYKIDK 749
QY 759 TYVYDHOIIPAAALRILGYFGITEKKLKASATGQKTLDF 797
Db 750 EYIDHQVLPVSLRILRFGYTEAQLKGAAB--QOTLDAF 787

RESULT 7

US-08-902-632-2
Sequence 2, Application US/08902632
Patent No. 6008025
GENERAL INFORMATION:
APPLICANT: KOMATSUBARA, Hideyuki
APPLICANT: KITABAVASHI, Masao
APPLICANT: KAMIMURA, Hideki
APPLICANT: KAWAKAMI, Bunsei
APPLICANT: KAWAMURA, Yoshihisa
APPLICANT: TAKAGI, Masahiro
APPLICANT: IWANAKA, Tadayuki
TITLE OF INVENTION: Modified Thermostable DNA Polymerase,
TITLE OF INVENTION: and DNA Polymerase Composition for Nucleic Acid
TITLE OF INVENTION: Amplification
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESS: Kenyon & Kenyon
STREET: 1 Broadway
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10004
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.50 inch floppy disk, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/902,632
FILING DATE: Concurrent Herewith

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 198911/96
FILING DATE: 29-JUL-1996
APPLICATION NUMBER: JP 200446/96
APPLICATION NUMBER: 30-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: GREASON, Edward W.
REGISTRATION NUMBER: 18,918
REFERENCE/DOCKET NUMBER: 2418/7
TELEPHONE: 212-425-7200
TELEFAX: 212-425-5288
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 774 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-902-632-2

Query Match 29.5%; Score 1237.5; DB 3; Length 774;
Best Local Similarity 36.6%; Pred. No. 3.6e-105;
Matches 303; Conservative 144; Mismatches 289; Indels 93; Gaps 22;

QY 8 VLDSSVGVGKEPQVILWIGIAENGERVVLIDRSFRPYFVALLAPGADPKQV---AQRIR 63
Db 2 ILDTYITEDGKPVIRIFK-KENGFEKIEYDRTFFPYFALLKDDSAIEEVKKITABRHG 60
QY 64 ALSRPKSPITIGVEDDKRYFGPRVLRIRTVLPEAVEYRELVKNVDPGVDELADIRF 123
Db 61 TVTVTKR----VEKQKKFLGRPEVWMLYFTHPDQVPAIRDKIREHGAVIDIYEVDIFF 116
QY 124 AMRYLIDHOLFPTFWYRVEAELENKMGFRVDKVLVKSRRPPLVGEALAPTKLPDLRL 183
Db 117 AKRYLIDKGLV-----PMEGD-----BELKML 138
QY 184 ADIEVYSKQSPRPDPVIVIAKTDGDEVL-----FLAEGKDDKPIREFVEY 235
Db 139 AFDITLYHEGEEFAB-GFILMSYADEEGARVITWKNVDLPYDVVSTEREMIKRFLRV 197
QY 236 VKRYDDIIVGVNNHFDWPIYLLRRARILGIKLDVTRVGABPTTSVHG--VSVPGRL 292
Db 198 VKEKDPDLITNGDNDFAYLKKRCEKLGINFALGRD-GSEPKIQRMGDRFAVEVKGR 256
QY 293 NVLDIYAEEMPEIKISLEAEVGVNKKSERVIVNWEIPDYDDPKRPLLLQYAR 352
Db 257 HFDLYPVIRRTINLPTTYLLEAVYEA VFGQPK-EKVYAE--EITPAWETGENLERVARYSM 313
QY 353 DVVRATYGLAEKILPAIQLSVYVGLPQGVGMSVGRLEWLYIRAFKMKELVFNRYE 412
Db 314 EDKATYELQKEFLPWEAQLSRIGSLNDVSRSSSTGNLVENWFLKAYERNELAPNKKPD 373
QY 413 RPE-----ETRGAVILEPLRGVHENTAVLDFSSMYNIMIKNVGPDILVRFGKCGC 467
Db 374 EKELARRRQSYEGYVKEPERGLWENIVYLDPRSLYPSIIITHNVSPDILNRE----- 426
QY 468 GCWE---APEVKHFRRCPCPGFKTVLERLLELRKVRAMKKYPPDPSPEYRLIDRQKA 524
Db 427 GCKEYDAPQVGHRCFKDPFGPIPSLLGLDLLEERQKIKKKMKA-TIDPTEKLLDYRQA 485
QY 525 LKVLANASYGVGMSGARWYCRCAKAVTANGRHILRTAI-NIARKLGLKVIYGDTSLSF 583
Db 486 IKILANSYGVYGYAPARWYCKEASVTANGREYITWTIKIEBKYGKVIYSDTSGF 545
QY 584 VTY---DPEKVE---NFIKIIEEL--GFEIKLEKVKYKELFFTEAKKRYAGLLEDGRID 634
Db 546 ATIPGADAETVKKAMEFLNYINAKLPGALEIEGYFKRGFFV-TKKYAVVIDESGKIT 604
QY 635 IVGFEAVRGDWCELAKEVQTKVVEIVLKTSEVNKAVYVRKIVKLEBEGKVPTEKLVINW 694
Db 605 TRGLEIVRRDWSIEIAKETQARVLEALLKDGDEKAVRIVKEVTEKLSKYEVPPEKLVHIE 664

QY 695 TSKRLBETTEAPHVAAKMLSGYRVSPGDKTGYVIVKGGGRISORAWPYFVMDPS 754
Db 665 QTRDLKDYKATGPHVAAKMLSGYRVSPGDKTGYVIVKGGGRISORAWPYFVMDPS 723
QY 755 --QIDVYVVDHQAALILGFGITEKLLKASATGOKTLDFELAKK 801
Db 724 KHKYDAEYIENQVLPVAVRILRATGFKEDLRYQKTRQVGLSAMLKPK 772

RESULT 8
US-08-688-649-37
; Sequence 37, Application US/08688649
; Patent No. 5827716
; GENERAL INFORMATION:
; APPLICANT: MAWONE, JOSEPH A.
; TITLE OF INVENTION: MODIFIED POL-II TYPE DNA
; TITLE OF INVENTION: POLYMERASES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; COUNTRY: California
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/688,649
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 220/281
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-688-649-37

Query Match 29.5%; Score 1237.5; DB 2; Length 776;
Best Local Similarity 36.1%; Pred. No. 3.6e-105;
Matches 301; Conservative 154; Mismatches 283; Indels 95; Gaps 23;

QY 8 VLDSSYVVGKPEQVITWGIANGERSVVLIDRSPPYFVYALLAPGADPKQV---AQRIR 63
Db 2 ILVDVYITEGKPVIRUFK-KENGKFKIEHDTFRPYFVYALLRDSKIEEVKTKTGRHG 60
QY 64 ALSRPKSPITIGVBDKRYKFRPRRVLIRTVLPEAVREYRELKVDGVDEVDLEADIRF 123
Db 61 KIVR-----IVDVKEKFKLGPITVMKLYLHEDQVFTREKVRHPAVVDIFEYDIPF 116
QY 124 AMRVLIHDHLPFTTWRVVEAPLENKMGFRVDKLYLVKSRPELYGSPALAPTLPDLRL 183
Db 117 AKRYLIDKGLI-----PMGE-----EELKIL 138
QY 184 AFDIEVTSKQSGPRPRDPVIVIAVKTDDGDEVL-----FIAEGKDDRKPIREFVEY 235

Db 139 AFDIETLYHGE-EFGKGPILMISYADENAEKVTWKVNDLPIYEVVSSREMIKFLRI 197
QY 236 VKRVDPIIIVYNNHEDWPLLRARILGILKDVTRVGAETTSVHG---VSVPGRL 292
Db 198 IREKDPDIIVYNGDSFDFYLAARAKLGIKLTIGRD-GSEPMQORIGDWTAVEVGRI 256
QY 293 NVLDYVAAEMPEIKIKSLEVAEYLGVMKKSRLVINWWEIPDYDDDPKRRLLIQLYAR 352
Db 257 HFDLYHVTITINLPTVYLEAVZAI-FGAPKEKVAD--EIAKESGGLNLRVAKYSM 313
QY 353 DDVRAATYGLAEKILPFAIQLSYVYGLDQVGAMSGVFRLEWYLIRAAFRMKELVNRVE 412
Db 314 EDAKATYELGKELPWEIQLSRVGLDQVPLDVSSTGNLWENFLLKAKARNEVANKPS 373
QY 413 RPE-----ETRGAVLEPLRGVHENIAVDFSSMYPNIMIKYVGPDTLVRPBGKCGE 466
Db 374 EEEYQRLRESYTGFGVKEPEKGLWENIVLDFRALYPSIIITHMVSPDTNLGE--- 428
QY 467 CGCWE-APEVKHFRRCPPGFEKTVLERILLELRKVRRAEMKKYPPDSPRYRLDEROKAL 525
Db 429 CKYVDIAPQVGHKFDKIDFIPSLCHLLEERQKIKTME--TQDPIEKILLDYRQKAI 487
QY 526 KVLNASYGYNGMSGARWYCRECAKAVTAWGRHLIRTA-INTARKLGLKVIYDGTOSLFV 584
Db 488 KLLANSFYGYGYAKARWYCKEACASVTANGRYVIELVMKLEBEKFGKVLIDTDLVA 547
QY 585 TYD-----PEKVENFIKIKELG--PEIKLEKYTKLFFTEAKRYAGLLEGRIDI 635
Db 548 TIFGSESEETKKALEFVKYINSKLPGLLESEYGFYKRGFFV-TKKRYAVIDEKGVIT 606
QY 636 VGFEAVRGDMCELAKEVOTKQVVEIVLKTSEVNAKAVYVRKIVKLEBEKGVPIEKLVIWKT 695
Db 607 RGLIEVRDWESEIAKTQARVLETILKHGDEEVRIVKEVIOKLANYSIPPEKLAISQ 666
QY 696 LSKRLEEYTEAPHVAAKMLSGYRVSPGDKTGYVIVKGGGRISORAWPYFVMDPS 753
Db 667 ITRPLEYKAIGPHVAAKMLSGYRVSPGDKTGYVIVKGGGRISORAWPYFVMDPS 723
QY 754 S--QIDVYVVDHQAALILGFGITEKLLKASATGOKTLDFELAKK 803
Db 724 KHKYDAEYIENQVLPVAVRILRATGFKEDLRYQKTRQVGLSAMLKPK 772

RESULT 9
US-07-966-278-1
; Sequence 1, Application US/07966278
; Patent No. 5483523
; GENERAL INFORMATION:
; APPLICANT: Mathur, Eric A
; TITLE OF INVENTION: EXONUCLEASE-DEFICIENT THERMOSTABLE
; TITLE OF INVENTION: PYROCOCUS FURIOSUS DNA POLYMERASE I
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thomas Fitting
; STREET: 12526 High Bluff Road, Suite 300
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92130
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/966,278
; FILING DATE: 19921226
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

NAME: Fitting, Thomas
 REGISTRATION NUMBER: 34,163
 REFERENCE/DOCKET NUMBER: STG0133P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-792-3680
 TELEFAX: 619-792-8477
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 775 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 US-07-966-278-1

Query Match 29.5%; Score 1236.5; DB 1; Length 775;
 Best Local Similarity 36.2%; Pred. No. 4.4e-105;
 Matches 301; Conservative 153; Mismatches 283; Indels 95; Gaps 23;

QY 8 VLDSSYEVVKGEPQVILWGIAENGVRVLLIDRSRPFYFALLAPGADPKOV---AQRIR 63
 DB 2 ILVDVITEGKPVIRLFK-KENGKFKIEHDTFRPYIALLRDDSKIEVKKITGERHG 60
 QY 64 ALSRPSPIIGVEDDKRKYFGPRRVLRIITVLPFAVREYRELKVNVDGVEDVLEADIRF 123
 DB 61 KIVR----IVDVEKVEKFKLGPITVWKLYLHPQDVPITREKVRHPAVVDIFEYDIPF 116
 QY 124 AMRYLIDHDLFPFTWYRVEAEPLNKGKFRVDKYLKSRPEPLVGEALAPTCLDLRIL 183
 DB 117 AKRYLIDKGLT-----PMEGS-----BELKIL 138
 QY 184 AFDIEVSKSGSPRPDPVIVIAVKTDDGDEVL-----FIAEGKDDRKPIREPVEX 235
 DB 139 AFDIETLYHEGE-FFGKPIIMISYADENAKVITWKNIDLPVIEVSSSEKIMKRLRI 197
 QY 236 VKRYDDPIIYGNVNNHEDWPLLRRARILGKLDVTRRVGAEPSTTVHGH---VSVEGRL 292
 DB 198 IREKDPDIIVYNGDSFDPVYLAKEAKGLIKLTIGRD-GSEPQMQIGDMTAVEVKGRI 256
 QY 293 NVLDIYDAEMPEKIKSLEBVARLYGWMKSERVIINWHEIPDYDDPKRPLLOQYAR 352
 DB 257 HFDLHYVITRINPTLTLEAVYEA1-FGPKERKYAD--EIAKAWESGENLERVAKYSM 313
 QY 353 DDVRATYGLABKILPFAIQLSYVTGLPDDVGAMSGVFRLEWYLIRAAEFMKELVNRVE 412
 DB 314 EDATYELGKEFLPMELQISRLVQPLWDSRSTGNLVEFWLLRKAYERNEVAPNKPS 373
 QY 413 RPE-----ETRYGAILVLEPLRGVHENIAVLDFSSMYPNIMIKYNGPDTLVRPGEKCGE 466
 DB 374 EEEYQRRLRESYTGFGVKEPEKGLWENIVLDFRPLPSIIITHNVSPDTINLEG-----428
 QY 467 CGCHE-APVVKHRRFCPPGFKTVLRELLEKRVRAEMKYPDPSPFRLDROKAL 525
 DB 429 CKNYDIAPOVGHKFCQIPGSPISLLGHLLEERQIKTKKE-TQDPIEKILLDYROKAI 487
 QY 526 KVLNAGSYGVNGWAGWYCRECAKAVTAWGRHLIRTA-INIARKGLKVITYGDTSLFV 584
 DB 488 KLLANSFYGYGYAKAWYCKEASWTAGWKYIELVWKEBEKFGKVLVYIDTDGLYA 547
 QY 585 TYD-----PEKVENFIKIIEBELG--FEIKLEKVKYRLFFTEAKRYAGLLEDGRIDI 635
 DB 548 TIPGSEBEIKKALEFVKYINSKLPGLLEVEYGFYKRGFFV-TKRYAVIDEGKVIT 606
 QY 636 VGPFAVRDNCWELAEQTKVVEIVLKTSEVKNKAVEYVRKIVKSLBEGKVPTEKLIWKT 695
 DB 607 RGLSIVRDMSEIAKETQARVLETILKHGDVEEAVRVKVEIQKLANYEIPPEKLAIEYQ 666
 QY 696 LSKLEEVTEAPHVAAKRLMAGYRVSPGDKIGYVIVKGGRIISORANPYFNVK--DP 753
 DB 667 ITRFHEKATGPHVAVAKLAAKGVKIKPGWVYIYVLRKGDGPISNRA---ILAEYDP 723
 QY 754 S--QIDVTYVVDHQIIPAALRIILGYFITEKKLASATGQKTLDFDL-AKKS 802

DB 724 KKHKYDAEYIENQVLPVLRILEGFGYRKEDLYQKTRQVGLTSWLNKKS 775
 RESULT 10
 US-08-424-921-1
 Sequence 1, Application US/08424921
 Patent No. 5545552
 GENERAL INFORMATION:
 APPLICANT: Machur, Eric A
 TITLE OF INVENTION: PURIFIED THERMOSTABLE PYROCOCUS
 TITLE OF INVENTION: FURIOSUS DNA POLYMERASE I
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Bingham & Fitting
 STREET: 12526 High Bluff Road, Suite 300
 CITY: San Diego
 STATE: CA
 COUNTRY: USA
 ZIP: 92130
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/424,921
 FILING DATE: 19-APR-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/803,627
 FILING DATE: 02-DEC-1991
 APPLICATION NUMBER: US 07/620,568
 FILING DATE: 03-DEC-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/657,073
 FILING DATE: 19-FEB-1991
 APPLICATION DATA:
 APPLICATION NUMBER: US 07/776,552
 FILING DATE: 15-OCT-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Fitting, Thomas
 REGISTRATION NUMBER: 34,163
 REFERENCE/DOCKET NUMBER: STG0100P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-792-3680
 TELEFAX: 619-792-8477
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 775 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 US-08-424-921-1

Query Match 29.5%; Score 1236.5; DB 1; Length 775;
 Best Local Similarity 36.2%; Pred. No. 4.4e-105;
 Matches 301; Conservative 153; Mismatches 283; Indels 95; Gaps 23;

QY 8 VLDSSYEVVKGEPQVILWGIAENGVRVLLIDRSRPFYFALLAPGADPKOV---AQRIR 63
 DB 2 ILVDVITEGKPVIRLFK-KENGKFKIEHDTFRPYIALLRDDSKIEVKKITGERHG 60
 QY 64 ALSRPSPIIGVEDDKRKYFGPRRVLRIITVLPFAVREYRELKVNVDGVEDVLEADIRF 123
 DB 61 KIVR----IVDVEKVEKFKLGPITVWKLYLHPQDVPITREKVRHPAVVDIFEYDIPF 116
 QY 124 AMRYLIDHDLFPFTWYRVEAEPLNKGKFRVDKYLKSRPEPLVGEALAPTCLDLRIL 183
 DB 117 AKRYLIDKGLT-----PMEGS-----BELKIL 138

```

QY 184 AFDIEVSKQSPRPDPVIVIAVKTDDGDEVL-----FIAEGKDRKPIREFVEY 235
Db 139 AFDIETLYHEGE-EFGKGLIMISYADENEAKVITWKNDLPYVEVSSEREMIKRFLRI 197
QY 236 VKRYDPIIIVGNNHFDWYLLRRARILGKLDVTRRVGABPTTSVHGH---VSVPGRL 292
Db 198 IREKDPDIIIVYNGSDSPDPPYLAKEAKLGKLTIGRD-GSEPKQRIQDWTAVEVGRI 256
QY 293 NVLDYDABEMPEIKISLEVAEYLGVMKKSERVIINWWEIPDYWDPPKRLQLQYAR 352
Db 257 HFDLVHVTITRTINLTPTTLEAVYEA-EGKPEKVVAD--EIAKAWESGENLERVAKYSM 313
QY 353 DVVRATYGLAEKILPPAIQLSVYTGCLPQVGMVSGFRLEWYLIIRAAFMKELVNRVE 412
Db 314 EDKATYELGKEFLPWEIQLSLVGOPLWDVSRSTGNLVWFLRKAYERNEVAPNPKS 373
QY 413 RPE-----ETRGALVLEPLRGVHENIAVLDFSSWYPMIMIKYVNPDDTLVRPCKGCE 466
Db 374 EEEYQRRRESYTGFGVKEPEKGLWENIYLDPRALYPSIIITHNVSDTLNLEG----- 428
QY 467 CGCWE-APEVKHFRRCPPGFFKTVLERLLELRKVRABMKYPPDPSYRLLDERQKAL 525
Db 429 CKNYDIAQVGHKFKCDIPGFTPSLLGHLEBRQKIKTKWE-TQDPIEKILLDYRQKAI 487
QY 526 KVLNASYGYMGWSGARWYCRCAKAVTANGRLHRTA-INTARKLGLKVIYGDTSDFV 584
Db 488 KLLANSFYGYGYAKARWYCKEASVTANGRYIELVWKELEBEKFGKVLVIDTDLGYA 547
QY 585 TYD-----PEKVENFIKIEELG--FEIKLEKVKYKLFTEAKKRYAGLLEDGRDI 635
Db 548 TTPGSEBIEKKALEFVKYIINSKLPGLULEYEGFYKRGFPV-TKKRYAVIDBEGKVIT 606
QY 636 VGFEAVRGDWCALAKEVOTKVVEIVLKTSEVNVKAYEVYVKIVKELEEGKVPEKIVWKT 695
Db 607 RGLVIRVRODSIAKETOARVLETILKHGDVEAVYIVKEVQKLANVEIPEKLAIVEQ 666
QY 696 LSKRLVEYTEAPHVAAKRLMSAGYVSPGDKIGYIVKGGGRISQRAWPYFMVK--DP 753
Db 667 ITRPLEHYKAIGHVAVAKLAAGVKIKPGWVIGYIVLRGDPISNEA---ILAEVDP 723
QY 754 S-QJDTVTVYVDHIIIPALRIILGYFGITEKLLKASATGKTLDFL-AKKS 802
Db 724 KKHGDAEYIENQVLPVLRILEGFGYRKEDLRYQKTRQVGLTSWLNIRKS 775

RESULT 11
US-08-556-355A-1
; Sequence 1, Application US/08556355A
; Patent No. 5866395
; GENERAL INFORMATION:
; APPLICANT: MATHUR, Eric J.
; TITLE OF INVENTION: Purified Thermostable Pyrococcus
; TITLE OF INVENTION: furiosus DNA Polymerase I
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSES: Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/556.355A
; FILING DATE: 13-NOV-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/424,921

```

```

; FILING DATE: 19-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/803,627
; FILING DATE: 02-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/779,846
; FILING DATE: 21-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,073
; FILING DATE: 19-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/620,568
; FILING DATE: 03-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Barker, M. Paul
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 04121.0004-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 775 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-556-355A-1

Query Match 29.5%; Score 1236.5; DB 2; Length 775;
Best Local Similarity 36.2%; Pred. No. 4.4e-105;
Matches 301; Conservative 153; Mismatches 283; Indels 95; Gaps 23;

QY 8 VLSSIEVCKEQQVINGTAENGERRVLLIDRSFRPYFALLAPGADPKQV----AQRIR 63
Db 2 ILVDVYTERGKPVIRLFK-KENGPKIEHDTFRPYIALLRDDSKIEVKKITGERHG 60
QY 64 ALSRPSFIIGDEDDKRYGPRRVLRTINTLPEAVREYRELAKNVGDGVLEADIRF 123
Db 61 KIVR---IVDVKEKFKLGKPIITWKLYLEHPQDPTIREKRHPAVVDIFEYDIF 116
QY 124 AMRYLDHDLFPFTWYRVEAEPLNKMGRVDKYLKSRPEPLYGEALAPTKLPDLRI 183
Db 117 AKRYLDKGLI-----PMEGE-----EBKIL 138
QY 184 ARLIEVSKQSPRPDPVIVIAVKTDDGDEVL-----FIAEGKDRKPIREFVEY 235
Db 139 AFDIETLYHEGE-EFGKGLIMISYADENEAKVITWKNDLPYVEVSSEREMIKRFLRI 197
QY 236 VKRYDPIIIVGNNHFDWYLLRRARILGKLDVTRRVGABPTTSVHGH---VSVPGRL 292
Db 198 IREKDPDIIIVYNGSDSPDPPYLAKEAKLGKLTIGRD-GSEPKQRIQDWTAVEVGRI 256
QY 293 NVLDYDABEMPEIKISLEVAEYLGVMKKSERVIINWWEIPDYWDPPKRLQLQYAR 352
Db 257 HFDLVHVTITRTINLTPTTLEAVYEA-EGKPEKVVAD--EIAKAWESGENLERVAKYSM 313
QY 353 DVVRATYGLAEKILPPAIQLSVYTGCLPQVGMVSGFRLEWYLIIRAAFMKELVNRVE 412
Db 314 EDKATYELGKEFLPWEIQLSLVGOPLWDVSRSTGNLVWFLRKAYERNEVAPNPKS 373
QY 413 RPE-----ETRGALVLEPLRGVHENIAVLDFSSWYPMIMIKYVNPDDTLVRPCKGCE 466
Db 374 EEEYQRRRESYTGFGVKEPEKGLWENIYLDPRALYPSIIITHNVSDTLNLEG----- 428
QY 467 CGCWE-APEVKHFRRCPPGFFKTVLERLLELRKVRABMKYPPDPSYRLLDERQKAL 525
Db 429 CKNYDIAQVGHKFKCDIPGFTPSLLGHLEBRQKIKTKWE-TQDPIEKILLDYRQKAI 487
QY 526 KVLNASYGYMGWSGARWYCRCAKAVTANGRLHRTA-INTARKLGLKVIYGDTSDFV 584
Db 488 KLLANSFYGYGYAKARWYCKEASVTANGRYIELVWKELEBEKFGKVLVIDTDLGYA 547

```


Db 257 HFDLYPVIRITINLPTTLEAVVEAVFGPK-EKVYAE--BITPAWETGENTLERVARYSM 313
QY 353 DDVRATYGLAEKILPFAIQLSYVTGLDQVGAMSGFRLEWYLIRAAFKMKELVNRVE 412
Db 314 EDKAVTYELGKEFLPMEQSLRLIGQSLMDVSRSSGNLVENWELLKAYERNELAPNKPD 373
QY 413 RPE-----ETVRGAIVLEPLRGVHENTAVLDFSSMYPNIMIKYNVGPDTLVRPGEKCGEC 467
Db 374 EKELARRRQSYEGGYVKEPGRGLWENTVYLDFFSLYPSIIITHNVSPDTLNRE----- 426
QY 468 GCWE--APEVKHPRCPGPFKTVLERLLELKRVRARMKKVPPDSPEYRLDEROKA 524
Db 427 GCKEYDVAPQVGHFCKDPGFPISLGDLEERQIKKQKKA-TIDPIERKLLDYRQRA 485
QY 525 LKVLANSYGYMGWSGARWYCRBCAKAVTAWGRHLINTAI-NIARKLGLKVITYGDTDSLF 583
Db 486 IKILANSYGYGYVARARWYCKEACSVTAWGREYITMTIKETEEKYGFVKVIYSDTDGFF 545
QY 584 VTY---DPEKYE---NFIKIKBEL--GFEIKLEKYKELFTEAKRYAGLLEGRID 634
Db 546 ATIPGADAEYVKKAMEFLNTYNIAKLPGALEYEGYKRGFFV-TKKYAVIDEKGIT 604
QY 635 IVGFEAVRGDWCELAKVEQTKVBEIVLKTSEVANKAVEYVRKIVKELEEGKVPTEKLVIMK 694
Db 605 TRGLEIVERDWESEIAKETQARVLEALLKGDGDEKAVRIVKEVTEKLSKYEVPEPEKLVHE 664
QY 695 TSKLELBEYTTTEAPHVNAKEMLSAGYRVSPGDKIGYVIVKGGGRISQRAWPYFMVKDPS 754
Db 665 QITRDLKDYKATGPHVAVAKELAARGVKIRPGTVISIVILKSGSRIGDRAIP-FDEFDPT 723
QY 755 --QIDVTYVDHQIIPAALRILGYFGITEKRLKASATGOKTLDFDLAKK 801
Db 724 KHKYDABYYIENQVLPAPERILRAFYGKEDDLRYOKTRQVLSANLKP 772

Search completed: November 25, 2003, 15:00:59
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2003, 12:04:39 ; Search time 8726 Seconds
(without alignments)
11308.047 Million cell updates/sec

Title: US-10-034-849-1

Perfect score: 2412

Sequence: 1 atgactgaagtgtatttcac.....tagccaagaagcaagtaa 2412

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2412	100.0	2412	6	AR072898	Sequence
2	2412	100.0	2412	6	AX469421	Sequence
3	2412	100.0	2412	6	AX469431	Sequence
4	2412	100.0	2412	6	BD080287	Isolation
5	2400.8	99.5	2412	6	AX469433	Sequence
6	1071.2	44.4	3068	6	E10436	DNA encoding
7	1071.2	44.4	3069	1	PVRDNPAB	Pyrodictium
8	1069.6	44.3	2430	6	I17571	Sequence 3
9	1045.6	43.3	2430	6	I17570	Sequence 1
10	822.2	34.1	185300	1	AP000063	Aeropyrum
11	735.2	30.5	2319	1	AB017501	Aeropyrum
12	735.2	30.5	2319	6	E36726	Novel DNA p
13	580	24.0	2610	6	AX035158	Sequence
14	580	24.0	3500	1	AF195019	Pyrobacul
15	574.6	23.8	266050	1	AP000988	Sulfolobu
16	570	23.6	10560	1	AE009857	Pyrobacul
17	569.8	23.6	2346	1	AB032376	Sulfurisp
18	326	13.5	12971	1	AE001070	Archaeogl
19	279	11.6	2367	6	AR072899	Sequence
20	279	11.6	2367	6	AX469423	Sequence
21	279	11.6	2367	6	BD080288	Isolation
22	204.2	8.5	2376	6	E13954	Artificial
23	200.8	8.3	3731	6	E13952	Thermococu
24	191.4	7.9	2325	6	E15057	gDNA encodi
25	191.4	7.9	2325	6	E15069	gDNA encode
26	191.4	7.9	2325	6	E15085	gDNA encodi
27	189.8	7.9	2325	6	BD175553	Modified
28	189.6	7.9	2325	6	A93613	Sequence 4
29	184.8	7.7	2325	6	BD078736	B type DN
30	177.4	7.4	2322	6	AX411312	Sequence
31	174.8	7.2	2331	6	AX135456	Sequence
32	174.4	7.2	2374	6	E14136	DNA encoding
33	174.4	7.2	2374	6	E14137	DNA encoding
34	173	7.2	2840	1	SSU92874	Sulfolobus
35	167.2	6.9	265118	1	CNSPAX06	Pyrococcu
36	167.2	6.9	265118	6	AX041922	Sequence
37	163.4	6.8	2995	1	PAPABPOL	P. abyssal
38	163.2	6.8	2322	6	A79152	Sequence 6
39	163.2	6.8	2322	6	BD009950	Thermococ
40	162.6	6.7	4446	1	PSP710POL	Pyrococcus
41	162.6	6.7	4446	6	A68743	Sequence 1
42	162.2	6.7	2995	6	A68744	Sequence 2
43	157.4	6.5	2408	1	TSU47108	Thermococu
44	150.8	6.3	2391	1	PSP250332	Pyrococcu
45	146.2	6.1	2734	6	I35626	Sequence 13

ALIGNMENTS

RESULT 1
AR072898

LOCUS

DEFINITION AR072898

ACCESSION AR072898

VERSION AR072898.1

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE

1 (bases 1 to 2412)

AUTHORS Callen, W. and Mathur, E.J.

TITLE Isolation and identification of polymerases

JOURNAL Patent: US 5948666-A 3 07-SEP-1999;

FEATURES Location/Qualifiers

AR072898 Sequence 3 from patent US 5948666. 2412 bp DNA linear PAT 28-AUG-2000

source		1. .2412																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		</	
--------	--	----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----	--

[illegible]

Db 601 GATCCTGTAATAGTAGTAGCTGTGAAGACTGACGATGGCGATGAGGTGCTATTCAATTGCA 660
QY 661 GAGGCGAAGACGATCGAAGACCGATACCGGAGTTGTAGAGTACGTGAAGAGGTATGAC 720
Db 661 GAGGCGAAGACGATCGAAGACCGATACCGGAGTTGTAGAGTACGTGAAGAGGTATGAC 720
QY 721 CCGGACATATAGTCGGGTATTAACAACAATCAATTCGATTGGGCTTATCTTTTGGAGCGC 780
Db 721 CCGGACATATAGTCGGGTATTAACAACAATCAATTCGATTGGGCTTATCTTTTGGAGCGC 780
QY 781 GCGCGATCCTAGGATTAAGCTTGTATGATGCTAGAGAGTGTGGCGCGAGCCACCACT 840
Db 781 GCGCGATCCTAGGATTAAGCTTGTATGATGCTAGAGAGTGTGGCGCGAGCCACCACT 840
QY 841 AGCTATCATGGGCAAGCTCTCTGCTCCCTGGCAGGCTTAACCTAGTCTGTACGACTATGCC 900
Db 841 AGCTATCATGGGCAAGCTCTCTGCTCCCTGGCAGGCTTAACCTAGTCTGTACGACTATGCC 900
QY 901 GAAGAGATGCCAGAGATCAAGATAAAGAGTCTCGAGAGCTCGCAGAGTATCTAGGCGTG 960
Db 901 GAAGAGATGCCAGAGATCAAGATAAAGAGTCTCGAGAGCTCGCAGAGTATCTAGGCGTG 960
QY 961 ATGAAGAGAGTGAACCGCTTATCATCAATTGGTGGAGATTCAGACTATTCGGACGAC 1020
Db 961 ATGAAGAGAGTGAACCGCTTATCATCAATTGGTGGAGATTCAGACTATTCGGACGAC 1020
QY 1021 CCGAAGAGAGACCACTATTACTGCAATACGCGCGCAGCATGTCCGCGTACTTACGGC 1080
Db 1021 CCGAAGAGAGACCACTATTACTGCAATACGCGCGCAGCATGTCCGCGTACTTACGGC 1080
QY 1081 TTAGCCGAGAGATATTGCGGTTGCTATCCAGTTGCTGACGATGACAGGTCTCCCACTA 1140
Db 1081 TTAGCCGAGAGATATTGCGGTTGCTATCCAGTTGCTGACGATGACAGGTCTCCCACTA 1140
QY 1141 GACCAGTAGTGCGATGAGTGTGGCTTTCGACTTGAATGGTACCTGATACGCGCGCG 1200
Db 1141 GACCAGTAGTGCGATGAGTGTGGCTTTCGACTTGAATGGTACCTGATACGCGCGCG 1200
QY 1201 TTTAAGATGAAGAGCTTGTGCCAAGCGGTTGACGCGCCGAGAGAGCTTACCGTGGC 1260
Db 1201 TTTAAGATGAAGAGCTTGTGCCAAGCGGTTGACGCGCCGAGAGAGCTTACCGTGGC 1260
QY 1261 GCTATAGTCTTTGAGCGGTTGAGAGGCGTGACAGAGATATAGCGTACTCGACTTTAGC 1320
Db 1261 GCTATAGTCTTTGAGCGGTTGAGAGGCGTGACAGAGATATAGCGTACTCGACTTTAGC 1320
QY 1321 TCGATGTACCCAAACATCATATGATAAAGTACAAATGTTGGTCTGTGACACGCTTGTAGGCGCT 1380
Db 1321 TCGATGTACCCAAACATCATATGATAAAGTACAAATGTTGGTCTGTGACACGCTTGTAGGCGCT 1380
QY 1381 GGTGAAAAGTGTGGAGTGTGTTGCTGGAGAGCCCGGAGGTCAAGCAAGGTTCCTGT 1440
Db 1381 GGTGAAAAGTGTGGAGTGTGTTGCTGGAGAGCCCGGAGGTCAAGCAAGGTTCCTGT 1440
QY 1441 AGGTGTCCGCGCGCTTCTTCAAGACAGCTTCTTCAAGAGCTTGTAGAGCTTCGTAAGCGT 1500
Db 1441 AGGTGTCCGCGCGCTTCTTCAAGACAGCTTCTTCAAGAGCTTGTAGAGCTTCGTAAGCGT 1500
QY 1501 GTGCGTGTGAAATGAAGAGTATCTCCGATAGCCAGAGATATCGACTTGTGGATGAA 1560
Db 1501 GTGCGTGTGAAATGAAGAGTATCTCCGATAGCCAGAGATATCGACTTGTGGATGAA 1560
QY 1561 AGGCAGAGGCGTGTGAAGTTCTTGCAACCGCTAGTACCGCTACATGGGTGAGCGGC 1620
Db 1561 AGGCAGAGGCGTGTGAAGTTCTTGCAACCGCTAGTACCGCTACATGGGTGAGCGGC 1620
QY 1621 GCTAGTGTGATTCGAGGAGTGTGCGAAGGCTGTACCGCTTGGGTAGGACCTCATA 1680
Db 1621 GCTAGTGTGATTCGAGGAGTGTGCGAAGGCTGTACCGCTTGGGTAGGACCTCATA 1680
QY 1681 CGCACCGCCATCAACATAGCTCGTAAACTAGGCGCTCAAGGTGATCTACGGTGAACAGAT 1740

Db 1681 CGCACCGCCATCAACATAGCTCGTAAACTAGGCGCTCAAGGTGATCTACGGTGAACAGAT 1740
QY 1741 TCGCTCTTCTGCTGACCTATGATCCGAGAGAGTGGAAAATTTTCATCAAAATTAAGAG 1800
Db 1741 TCGCTCTTCTGCTGACCTATGATCCGAGAGAGTGGAAAATTTTCATCAAAATTAAGAG 1800
QY 1801 GAGCTGGGTTTCGAAATCAAGCTAGAGAGGTGTGTAACACGCTTATCTTTACAGAGCT 1860
Db 1801 GAGCTGGGTTTCGAAATCAAGCTAGAGAGGTGTGTAACACGCTTATCTTTACAGAGCT 1860
QY 1861 AAGAAGAGTACGCTGGGCTTCTCGAGGACGAGTATAGATATTGTGCGTTTCAGAGCT 1920
Db 1861 AAGAAGAGTACGCTGGGCTTCTCGAGGACGAGTATAGATATTGTGCGTTTCAGAGCT 1920
QY 1921 GTACGTGGCGATTGGTGTGAATCCGCAAGAGGTTCAGACTAAGGTTGTGCAAAATAGTA 1980
Db 1921 GTACGTGGCGATTGGTGTGAATCCGCAAGAGGTTCAGACTAAGGTTGTGCAAAATAGTA 1980
QY 1981 TTGAAGACGAGTGAAGTGAACGAGCTGTAGAGTACGTCAGGAAGATTGTGAAGAGTTG 2040
Db 1981 TTGAAGACGAGTGAAGTGAACGAGCTGTAGAGTACGTCAGGAAGATTGTGAAGAGTTG 2040
QY 2041 GAGGAGGCAAGGTTCCCATAGAGAGCTTGTAACTCTGGAAGACCTTGTAGTAAAGCTTT 2100
Db 2041 GAGGAGGCAAGGTTCCCATAGAGAGCTTGTAACTCTGGAAGACCTTGTAGTAAAGCTTT 2100
QY 2101 GAGGAGTACACAAACGAGGACACACACGCTGTCAGGAGAGAGGATGCTGCAGCAGC 2160
Db 2101 GAGGAGTACACAAACGAGGACACACACGCTGTCAGGAGAGAGGATGCTGCAGCAGC 2160
QY 2161 TACCGGTTAAGCCCGAGGACAAAGATAGAGTATGTAATAGTGAAGGTTGGCGGTATC 2220
Db 2161 TACCGGTTAAGCCCGAGGACAAAGATAGAGTATGTAATAGTGAAGGTTGGCGGTATC 2220
QY 2221 AGTCAAGAGCATGSCCATACTTCTGTCAGAGGATCTTAGCCAGATAGACGTGACCTAC 2280
Db 2221 AGTCAAGAGCATGSCCATACTTCTGTCAGAGGATCTTAGCCAGATAGACGTGACCTAC 2280
QY 2281 TATGTTGACCAACCAATCATCCCGCTGANTTGAAGATCTGGGCTACTTTGGCATCAC 2340
Db 2281 TATGTTGACCAACCAATCATCCCGCTGANTTGAAGATCTGGGCTACTTTGGCATCAC 2340
QY 2341 GAGAAGAGCTGAAGAGCAAGTCAACTGGGCAAGAGTCTCTTCGACTTTTAGCCAAAG 2400
Db 2341 GAGAAGAGCTGAAGAGCAAGTCAACTGGGCAAGAGTCTCTTCGACTTTTAGCCAAAG 2400
QY 2401 AAGAGCAAGTAA 2412
Db 2401 AAGAGCAAGTAA 2412

RESULT 4

BD080287 LOCUS 2412 bp DNA linear PAT 27-AUG-2002
DEFINITION Isolation and identification of novel polymerases.

BD080287 ACCESSION

BD080287.1 GI:22625890

Pyrobolus fumarii

ORGANISM

Pyrobolus fumarii
Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
Pyrodictiaceae; Pyrobolus.

1 (bases 1 to 2412)

Callen, W. and Mathur, E.J.

Isolation and identification of novel polymerases

Patent: JP 2001512678-A 2 28-AUG-2001;

DIVERSA INC

OS Pyrobolus fumarius

PN JP 2001512678-A/2

PD 28-AUG-2001

PF 06-AUG-1998 JP 2000506322

PR 06-AUG-1997 US 08/907166

PI WALTER CALLEN, ERIC J MATHUR

PC C12N15/09, C07K16/40, C12N1/15, C12N1/19, C12N1/21, C12N5/10 PC
, C12N9/12, G01N33/53
PC GOIN33/53, GOIN33/566, C12N15/00, C12N5/00
CC Isolation and identification of novel polymerases FH Key

Location/Qualifiers
 CDS (1) . (2410) .
 FT

```

FEATURES             Location/Qualifiers
     source            1..2412
                        /organism="Pyrobolus fumarii"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:54252"
     BASE COUNT       631 a 512 c 707 g 562 t

```

Query Match	100.0%	Score	2412;	DB	6;	Length	2412;
Best Local Similarity	100.0%;	Prod.	No. 0;				
Matches	2412;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	ATGACTGAGGTGTGTTATTCACCGTTTATAGACTCTAGCTACGAGGTGTGTTGGTAAAGAGCCT	60				
DB	1	ATGACTGAGGTGTGTTATTCACCGTTTATAGACTCTAGCTACGAGGTGTGTTGGTAAAGAGCCT	60				
QY	61	CAGGTAAATCATATGGGGTATTCGTGAGAACCGGAGAGGGTAGTCCTCATTTGACAGGTCT	120				
DB	61	CAGGTAAATCATATGGGGTATTCGTGAGAACCGGAGAGGGTAGTCCTCATTTGACAGGTCT	120				
QY	121	TTTTCGCCCATCTACTTCTATGCGCTGCTTCACCGGCCCGGATCCTTAAGCAGGTAGACAA	180				
DB	121	TTTTCGCCCATCTACTTCTATGCGCTGCTTCACCGGCCCGGATCCTTAAGCAGGTAGACAA	180				
QY	181	CGTATTCGTGCAATTCAGTAGGCCAAAGAGCCCGATTTATAGGTGTAGAGATGACAAAGG	240				
DB	181	CGTATTCGTGCAATTCAGTAGGCCAAAGAGCCCGATTTATAGGTGTAGAGAGATGACAAAGG	240				
QY	241	AAGTACTTCGGAGGCCCTCTAGAGGTCTTACGTATTCGACCGGTCTACCCGGAGCTGTT	300				
DB	241	AAGTACTTCGGAGGCCCTCTAGAGGTCTTACGTATTCGACCGGTCTACCCGGAGCTGTT	300				
QY	301	AGCGAGTATCGCGAACTCTGTAAAGAACGTTGATGGTGTAGAGATGTTCTAGAGCGGAT	360				
DB	301	AGCGAGTATCGCGAACTCTGTAAAGAACGTTGATGGTGTAGAGATGTTCTAGAGCGGAT	360				
QY	361	ATACGCTTCGTATGGGCTATCTATAGATACAGATCTATTTCCCTTACCTGGTACCGT	420				
DB	361	ATACGCTTCGTATGGGCTATCTATAGATACAGATCTATTTCCCTTACCTGGTACCGT	420				
QY	421	GTAGAGGCTGAGCCCTCGAGAACAAAGATGGGCTCCGCTGTCGACAAAGTATACCTGGTT	480				
DB	421	GTAGAGGCTGAGCCCTCGAGAACAAAGATGGGCTCCGCTGTCGACAAAGTATACCTGGTT	480				
QY	481	AAGAGCAGCCCGAGCCACTTTATGGTAGGCTCTCGCAACCAACAGCTTCCCGATCTT	540				
DB	481	AAGAGCAGCCCGAGCCACTTTATGGTAGGCTCTCGCAACCAACAGCTTCCCGATCTT	540				
QY	541	AGGATATCGCGTTCGATATTCAGGTTTATAGCAGCAAGGTCGCGCGGTCCAGAGCGC	600				
DB	541	AGGATATCGCGTTCGATATTCAGGTTTATAGCAGCAAGGTCGCGCGGTCCAGAGCGC	600				
QY	601	GATCTGTAAATAGTATAGCTGTGAAGACTGACGATGGCGATGAGGTGCTATTCAATGCA	660				
DB	601	GATCTGTAAATAGTATAGCTGTGAAGACTGACGATGGCGATGAGGTGCTATTCAATGCA	660				
QY	661	GAGGCGAAGACGATCGAAACCGATACCGGAGTTGTAGGTACGTGAAGAGGTATGAC	720				
DB	661	GAGGCGAAGACGATCGAAACCGATACCGGAGTTGTAGGTACGTGAAGAGGTATGAC	720				
QY	721	CCCGACATATAGTCGGTTATACAAACAATCATTTCCGATTTGGCCTTATCTTTTGGCGCG	780				
DB	721	CCCGACATATAGTCGGTTATACAAACAATCATTTCCGATTTGGCCTTATCTTTTGGCGCG	780				
QY	781	GCCCGCATCTAGGCAATAAGCTTGATGTGACTAGAGAGTGTGGCGCGAGCCACCACT	840				
DB	781	GCCCGCATCTAGGCAATAAGCTTGATGTGACTAGAGAGTGTGGCGCGAGCCACCACT	840				

```
QY 1921 GTACGTGGCGATTGGTGTGAACCTGCCAAGAGAGGTTTCAGACTAAGGTTTGTTCGAAATAGTA 1980
Db 1921 GTACGTGGCGATTGGTGTGAACCTGCCAAGAGAGGTTTCAGACTAAGGTTTGTTCGAAATAGTA 1980
QY 1981 TTGAAGACGATGAGGTGGAACAAGGCTGTGATAGTACGTACAGGAAGATTGTGAAGAGTTG 2040
Db 1981 TTGAAGACGATGAGGTGGAACAAGGCTGTGATAGTACGTACAGGAAGATTGTGAAGAGTTG 2040
QY 2041 GAGGAGGCGAAGTTCCTCATAGAGAGCTTGTATCTGGAAGACCTTGTAGTAGCGCTTT 2100
Db 2041 GAGGAGGCGAAGTTCCTCATAGAGAGCTTGTATCTGGAAGACCTTGTAGTAGCGCTTT 2100
QY 2101 GAGGAGTACACAACGAGGACACACACGTCTGTCAGCGAAGAGGATGCTGTACAGAGGC 2160
Db 2101 GAGGAGTACACAACGAGGACACACACGTCTGTCAGCGAAGAGGATGCTGTGTACAGAGGC 2160
QY 2161 TACCGGTAAAGCCAGCGCAAGATAGGATGTATATAGTGAAGGTTGTGCGCGTATC 2220
Db 2161 TACCGGTAAAGCCAGCGCAAGATAGGATGTATATAGTGAAGGTTGTGCGCGTATC 2220
QY 2221 AGTCAAGAGATGGCCATCTTCTATGTCAGAGTCCCTAGCCAGATAGAGCTGACCTAC 2280
Db 2221 AGTCAAGAGATGGCCATCTTCTATGTCAGAGTCCCTAGCCAGATAGAGCTGACCTAC 2280
QY 2281 TATGTTGACACCAAAATCATCCCGGTGCAATGAGATACTGGGCTACTTTGGCATCACC 2340
Db 2281 TATGTTGACACCAAAATCATCCCGGTGCAATGAGATACTGGGCTACTTTGGCATCACC 2340
QY 2341 GAGAAGAGCTGAAGCAAGTGCACCTGGCAGAGACTCTCTTCGACTTTCTAGCCAAG 2400
Db 2341 GAGAAGAGCTGAAGCAAGTGCACCTGGCAGAGACTCTCTTCGACTTTCTAGCCAAG 2400
QY 2401 AAGAGCAAGTAA 2412
Db 2401 AAGAGCAAGTAA 2412

RESULT 5
AX469433
LOCUS AX469433 2412 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 15 from Patent WO0220735.
ACCESSION AX469433
VERSION AX469433.1 GI:21901720
KEYWORDS
SOURCE Pyrolobus fumarii
ORGANISM Pyrolobus fumarii
REFERENCE 1
AUTHORS Callen, W., Mathur, E. J. and Short, J. M.
TITLE Enzymes having high temperature polymerase activity and methods of use thereof
JOURNAL Patent: WO 0220735-A 15 14-MAR-2002;
DIVERSA CORPORATION (US)
FEATURES
source
1. 2412
/organism="Pyrolobus fumarii"
/mol_type="genomic DNA"
/db_xref="taxon:54252"
BASE COUNT 627 a 513 c 712 g 560 t
ORIGIN

Query Match 99.5%; Score 2400.8; DB 6; Length 2412;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2405; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGACTGAAGTTGATTACCGGTTTACGCTTCTAGCTACGAGGTTGTGGTAAAGACCT 60
Db 1 ATGACTGAAGTTGATTACCGGTTTACGCTTCTAGCTACGAGGTTGTGGTAAAGACCT 60
QY 61 CAGGTAATCATATGCGGTTATGCTGAGAACCGGAGAGGGTAGTCTCATTTGACAGGTCT 120
Db 61 CAGGTAATCATATGCGGTTATGCTGAGAACCGGAGAGGGTAGTCTCATTTGACAGGTCT 120
```

```
QY 121 TTTCCGCCATCTTCTATGCGCTGTGCAACCGGCGCGGATCTTAAGCAGGTAGCACAA 180
Db 121 TTTCCGCCATCTTCTATGCGCTGTGCAACCGGCGCGGATCTTAAGCAGGTAGCACAA 180
QY 181 CGTATTCGTGCTANTGAGTAGGCCAAAGAGCCCGATTATAGGTGTAGAGGATGACAAGAGG 240
Db 181 CGTATTCGTGCTANTGAGTAGGCCAAAGAGCCCGATTATAGGTGTAGAGGATGACAAGAGG 240
QY 241 AAGTACTTCGGAGGACCTCGTAGGGTCTTAGGTCTTAGTATTCGACCGGTCTACCGAGGCTGTT 300
Db 241 AAGTACTTCGGAGGACCTCGTAGGGTCTTAGGTCTTAGTATTCGACCGGTCTACCGAGGCTGTT 300
QY 301 AGGGAGTATCGCGAACTCGTAAAGAACTGTGATGTGTGAGGATGTTCTAGAGGCGGAT 360
Db 301 AGGGAGTATCGCGAACTCGTAAAGAACTGTGATGTGTGAGGATGTTCTAGAGGCGGAT 360
QY 361 ATACGCTTCGCTATGCGCTATCTCATAGATCAGATCTATTTTCCTTTCACCTGGTACCGT 420
Db 361 ATACGCTTCGCTATGCGCTATCTCATAGATCAGATCTATTTTCCTTTCACCTGGTACCGT 420
QY 421 GTAGAGGCTGAGCCCTCGAGAACAGATGGGCTTCCGTGTGCGACAGGTATACCTGTT 480
Db 421 GTAGAGGCTGAGCCCTCGAGAACAGATGGGCTTCCGTGTGCGACAGGTATACCTGTT 480
QY 481 AAGAGCAGCGCGAGGCCACTTTATGCTGAGGCTCTCGCACCAACCAAGCTTCCCGANTCTT 540
Db 481 AAGAGCAGCGCGAGGCCACTTTATGCTGAGGCTCTCGCACCAACCAAGCTTCCCGANTCTT 540
QY 541 AGGATACTCGCTTCGATATTTGAAGTTTATAGCAAGCAAGGTCGCCCGCTCCAGAGGCG 600
Db 541 AGGATACTCGCTTCGATATTTGAAGTTTATAGCAAGCAAGGTCGCCCGCTCCAGAGGCG 600
QY 601 GATCTCTGTAATGATAGTGTGAAGACTGACGATGGGATGCGGATGAGTCTTATTTGCA 660
Db 601 GATCTCTGTAATGATAGTGTGAAGACTGACGATGGGATGCGGATGAGTCTTATTTGCA 660
QY 661 GAGGCAAGACGATCGCAAAACGATACGCGAGTTTGTAGAGTACGTGAAGAGGTATGAC 720
Db 661 GAGGCAAGACGATCGCAAAACGATACGCGAGTTTGTAGAGTACGTGAAGAGGTATGAC 720
QY 721 CCCGACATAATAGTCGGTTTATAACAATATTCGATTGGCCTTATCTTTTGGAGGCG 780
Db 721 CCCGACATAATAGTCGGTTTATAACAATATTCGATTGGCCTTATCTTTTGGAGGCG 780
QY 781 GCGCGATCCTTAGGCATTAAGCTTCAATGACTAGAGATTGGCGCCGAGGCCACCACT 840
Db 781 GCGCGATCCTTAGGCATTAAGCTTCAATGACTAGAGATTGGCGCCGAGGCCACCACT 840
QY 841 AGCGTACATGGGACGCTCTCTCTCGTGGCAGGCTTAAAGTATCTGTACGACTATGCC 900
Db 841 AGCGTACATGGGACGCTCTCTCTCGTGGCAGGCTTAAAGTATCTGTACGACTATGCC 900
QY 901 GAAGAGATCCAGAGATCAAGATAAAGAGTCTCGAGGAGGTGCGAGAGTATCTAGGCGTG 960
Db 901 GAAGAGATCCAGAGATCAAGATAAAGAGTCTCGAGGAGGTGCGAGAGTATCTAGGCGTG 960
QY 961 ATGAAGAGAGTGAACGGGTTATCATCAATTGCTGGAGATTCGAGACTATTTGGAGCGAC 1020
Db 961 ATGAAGAGAGTGAACGGGTTATCATCAATTGCTGGAGATTCGAGACTATTTGGAGCGAC 1020
QY 1021 CCGAAGAGAGACCACTATTACTGCAATACGCGCGGACGATGTCGCGCTACTTACCGC 1080
Db 1021 CCGAAGAGAGACCACTATTACTGCAATACGCGCGGACGATGTCGCGCTACTTACCGC 1080
QY 1081 TTAGCCGAGAAGATTTGCGGTTTGTCTATCCAGTTGTGCTACGTAACAGGTCTCCCACTA 1140
Db 1081 TTAGCCGAGAAGATTTGCGGTTTGTCTATCCAGTTGTGCTACGTAACAGGTCTCCCACTA 1140
QY 1141 GACCAGGTAGGTGCGATGAGTGTGGCTTTGCACTTGAATGATACCTGATACGCGCGG 1200
Db 1141 GACCAGGTAGGTGCGATGAGTGTGGCTTTGCACTTGAATGATACCTGATACGCGCGG 1200
```



```
QY 1201 TTTAGATGAAGAGCTGTGCGGAACCGCGTTGAGCGCCGCAAGAGACTTACCGTGGC 1260
Db 1201 TTTAAGTGAAGAGCTGTGCGGAACCGCGTTGAGCGCCGCAAGAGACTTACCGTGGC 1260
QY 1261 GCTATAGTCTTGGAGCGTTGAGAGCGGTGACAGAGATATAGCCGCTACTCGACTTTAGC 1320
Db 1261 GCTATAGTCTTGGAGCGTTGAGAGCGGTGACAGAGATATAGCCGCTACTCGACTTTAGC 1320
QY 1321 TCGATGTACCCAAACATCATGATATAGTACAATGTTGGTCTGACACGCTTGTGAGCGCT 1380
Db 1321 TCGATGTACCCAAACATCATGATATAGTACAATGTTGGTCTGACACGCTTGTGAGCGCT 1380
QY 1381 GGTGAAGAAGTGTGCGAGTGTGTTGCTGGAGGCCCCCGAGGTCAAGCACAGGTTCCGT 1440
Db 1381 GGTGAAGAAGTGTGCGAGTGTGTTGCTGGAGGCCCCCGAGGTCAAGCACAGGTTCCGT 1440
QY 1441 AGGTGTCGGCCCGCTTCTTCAAGACAGTCTTGGAGGCTGTAGAGCTTCGTAGCGGT 1500
Db 1441 AGGTGTCGGCCCGCTTCTTCAAGACAGTCTTGGAGGCTGTAGAGCTTCGTAGCGGT 1500
QY 1501 GTGCGTCTGAATGAAGAAGTATCTTCCGATAGCCAGAAATATCGACTGTTGGATGAA 1560
Db 1501 GTGCGTCTGAATGAAGAAGTATCTTCCGATAGCCAGAAATATCGACTGTTGGATGAA 1560
QY 1561 AGGCAGAGCGGTTGAAGTCTTTCGAAACGCTAGTTACCGCTACATGGGTTGGAGCGGC 1620
Db 1561 AGGCAGAGCGGTTGAAGTCTTTCGAAACGCTAGTTACCGCTACATGGGTTGGAGCGGC 1620
QY 1621 GCTAGTGTGTTGCGAGGAGTCCGCAAGGCTGTACGCGCTGGGGTAGGCACTCATA 1680
Db 1621 GCTAGTGTGTTGCGAGGAGTCCGCAAGGCTGTACGCGCTGGGGTAGGCACTCATA 1680
QY 1681 CGCACCCGCCATCAACATAGCTCGTAAACCTAGGCTCAAGGTGATCTACCGGTGACACAGAT 1740
Db 1681 CGCACCCGCCATCAACATAGCTCGTAAACCTAGGCTCAAGGTGATCTACCGGTGACACAGAT 1740
QY 1741 TCGCTCTTCGTGACCTATGATCCGAGAGGCTGACGAGTGAATTTCACTAAATATTAAGGAG 1800
Db 1741 TCGCTCTTCGTGACCTATGATCCGAGAGGCTGACGAGTGAATTTCACTAAATATTAAGGAG 1800
QY 1801 GAGCTGGGTTGCGAAATCAACAGCTAGAGAGGCTGACAAACGCTTATTTCTTACAGAGCT 1860
Db 1801 GAGCTGGGTTGCGAAATCAACAGCTAGAGAGGCTGACAAACGCTTATTTCTTACAGAGCT 1860
QY 1861 AAGAAGAGTACGCTGCGCTTCTCGAGGAGCGGATAGATATTTGCGGTTTCGAGGCT 1920
Db 1861 AAGAAGAGTACGCTGCGCTTCTCGAGGAGCGGATAGATATTTGCGGTTTCGAGGCT 1920
QY 1921 GTACGTGGCGATTGCTGTAACCTCCCAAGAGGCTCAGACTAAGGTTGTCGAAATAGTA 1980
Db 1921 GTACGTGGCGATTGCTGTAACCTCCCAAGAGGCTCAGACTAAGGTTGTCGAAATAGTA 1980
QY 1981 TTGAAGACGAGTGAAGTGAACAGGCTGTAGTACGTACGAGAGATTTGTGAAAGAGTTG 2040
Db 1981 TTGAAGACGAGTGAAGTGAACAGGCTGTAGTACGTACGAGAGATTTGTGAAAGAGTTG 2040
QY 2041 GAGGAGGCGAAGTTCCCATAGAGAGCTTGTAACTCTGGAAGACCCCTTAGTAAAGCGCTT 2100
Db 2041 GAGGAGGCGAAGTTCCCATAGAGAGCTTGTAACTCTGGAAGACCCCTTAGTAAAGCGCTT 2100
QY 2101 GAGAGTATACAAACGAGGAGCCACAGCTGCTTCCAGCAAGAGAGGCTGTGACGAGGC 2160
Db 2101 GAGAGTATACAAACGAGGAGCCACAGCTGCTTCCAGCAAGAGAGGCTGTGACGAGGC 2160
QY 2161 TACCGGTAAAGCCAGGCGCAAGATAGGCTATGTAATAGTGAAGGTTGGTGGCGGTATC 2220
Db 2161 TACCGGTAAAGCCAGGCGCAAGATAGGCTATGTAATAGTGAAGGTTGGTGGCGGTATC 2220
QY 2221 AGTCAAGAGCATGGCCATCTTCTATGCTCAAGGATCTTACGCCAGATAGAGCTGACCTAC 2280
Db 2221 AGTCAAGAGCATGGCCATCTTCTATGCTCAAGGATCTTACGCCAGATAGAGCTGACCTAC 2280
QY 2281 TATGTTGACCAACCAATCATCCCGGCTGCAATGAGAAATACCTGGGCTACTTTGGCATCAC 2340
```

```
Db 2281 TATGTTGACCAACCAATCATCCCGGCTGCAATGAGAAATACCTGGGCTACTTTGGCATCAC 2340
QY 2341 GAGAGAGAGCTGAAAGCAAGTGCACCTGGGCGAGAGACTCTCTTCGACTTCTAGCCAAAG 2400
Db 2341 GAGAGAGAGCTGAAAGCAAGTGCACCTGGGCGAGAGACTCTCTTCGACTTCTAGCCAAAG 2400
QY 2401 AAGAGCAAGTAA 2412
Db 2401 AAGAGCAAGTAA 2412

RESULT 6
LOCUS E10436 3068 bp DNA linear PAT 29-SEP-1997
DEFINITION DNA encoding DNA polymerase.
ACCESSION E10436
VERSION E10436.1 GI:22027269
KEYWORDS JP 1995327684-A/3.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 3068)
AUTHORS Uemori,T., Ishino,Y. and Katou,I.
TITLE DNA POLYMERASE GENE
JOURNAL Patent: JP 1995327684-A 3 19-DEC-1995;
TAKARA SHUZO CO LTD
COMMENT OS Pyrodicticum occultum
PN JP 1995327684-A/3
PD 19-DEC-1995
PF 03-JUN-1994 JP 1994150591
PI UEMORI TAKASHI, ISHINO YOSHIZUMI, KATOU IKUNOSHIN PC
C12N15/09,C12N9/12,(C12N15/09,C12N1/01),(C12N9/12,C12N1/19); CC
strandedness: Double;
CC topology: Linear;
PH Key Location/Qualifiers
PH source 1..3068 /organism='Pyrodicticum occultum' FT
FT /strain='DSM2709T' 276..2687
FT CDS 276..2687 /product='DNA polymerase'.
FEATURES
source
Location/Qualifiers
1..3068
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'
BASE COUNT 730 a 745 c 980 g 613 t
ORIGIN

Query Match 44.4%; Score 1071.2; DB 6; Length 3068;
Best Local Similarity 66.5%; Pred. No. 2.5e-286;
Matches 1602; Conservative 0; Mismatches 788; Indels 18; Gaps 4;

QY 11 TTGTATTCACGGTTTGTAGACTCTAGCTACGAGGTGTTGGTAAAGAGCTCAGGTAATCA 70
Db 289 TAGAGTTCGTGCTGTAGACTCTAGCTACGAGATCTGGGGAAGGCGGTAGTAATCC 348
QY 71 TATGGGTATGCTGAGAACGGCGAGAGGTAGTCTCATTCACAGAGTCTTTTCGCCCAT 130
Db 349 TCTGGGGATACGCTTGACGGTAAACGTGTGCTCTTAGACCCACCGCTTCGCCCT 408
QY 131 ACTTCTATCGCTGTGTGACCGGGCGCGATCCCTA-----AGAGGTAGCACAACTGA 184
Db 409 ACTTCTACGCCCTCATAGCCCGGGCTATGAGGATATGTTGGAGGAGATAGCAGCTTCCA 468
QY 185 TTCGTGCATGTAGTAGCCCAAGAGCCCGATTTATAGGTGTAGAGATGACAGAGGAGT 244
Db 469 TAAGGAGGCTTAGTGTGGTCAAGAGTCCGATTAATAGTCCCAAGGCTCTTGATAAGGT 528
QY 245 ACTTCCGGAGGCTCGTGTAGGCTTACGCTATTTCGACCGCTGTACCCGAGGCTGTTAGG 304
Db 529 ACTTCCGGAGGCTCGGCTAAGGGCGGTGAAGATTACCACTATGATACCCGAGTCTGTAGAC 588
```


Qy	305	AGTATCCGGAACCTCGTAAAGACGTTGATGGTGTGGATGTCTAGAGGCGGATATAC	364
Dd	589	ACTACCGGAGGGGTGAAGAAGATAGAGGTGTGGAGGACTCCCTCGAGCGAGATATA	648
Qy	365	GCTTCGTATGCGCTACTCATAGATCAGCATCTATTTCCTTTCACTGGTACCGGTAG	424
Dd	649	GGTTTGAATAGAGATCTCATAGATAAGAGGCTCTACCCGTTTCAAGGTTTACCGGATCC	708
Qy	425	AGGCTGAGCCCTCGAAGACAAGATGGGCTTCCGTGTCGACAAAGGTATACCTGTAAAGA	484
Dd	709	CCGTAGAGGATGCGGGCGCAATCCAGGCTTCCGTGTGACCGTGTCTACAAGGTTGCTG	768
Qy	485	GCAGGCGGAGCCATCTTTATGTTGAGGCTCTCGCACAACCAAGCTTCCGATCTTAGGA	544
Dd	769	GCAGCCCGAGCCCT-----AGCGATATAACGCGGATCGACCTTCCCGCATGAGGC	822
Qy	545	TACTCGGTTGCTATTTGAAGTTTATAGCAAGCAAGGTCGCGCGTCCAGAGCGGATC	604
Dd	823	TGGTAGCTTTTGATATAGAGGTGTATAGCAGGAGGGGAGCCCTTAACCTTGCAAGGATC	882
Qy	605	CTGTAAATAGTATAGCTGTGAAGACTACGATGCGGATGAGGTGCTATTCTATTCAGAGG	664
Dd	883	CAGTGATAATAGTCTCCCTGAGGACAGCGAGGCGAAGGAGGCTCATAGAGCTGAAG	942
Qy	665	GCAAGACGATCGGAAACCGATACGCGAGTTGTAGAGTACGTGAGAGGATGATGACCCG	724
Dd	943	GCCATGACGACAGGAGGTTCTGAGGAGTTCTGAGAGTACGTGAGAGCTTCCGACCCG	1002
Qy	725	ACATAATAGTCGGTTATACACACATCATTTGATTTGGCTTATCTTTGAGGCGGCC	784
Dd	1003	ACATAATAGTGGGCTATAACAGTAAACCACTTCGACTGGCCCTACCTTAATGAGCGGCC	1062
Qy	785	GCATCTTAGGCATAAAGCTTGATGTGACTAGAGAGTTGGCGCGAGCCCAACCACTAGCG	844
Dd	1063	GTAGGCTCGGATTAAGCTCGACGTTACACGCGTGTGGGGCAGAGCCCAACCCAGCG	1122
Qy	845	TACATGGGACGCTCTGTCCCTGGGAGGCTTAACGTAGATCTGTAGACTATGCGCGAAG	904
Dd	1123	TCTACGCCACGCTCTCGGTGAGGTTAGGCTGTAACGTGGACCTCTACGACTATGCGGAG	1182
Qy	905	AGATGCCAGAGATCAAGATAAAGATCTCGAGAGGTCGACAGTATCTAGGCGTGATGA	964
Dd	1183	AGATCCGAGATTAAGTTAGAGCGCTTGAGGAGGTAGCGAGTACTTAGGCGTTATGA	1242
Qy	965	AGAGAGTGAAACGGTTTATCATCAATTTGTTGGAGATTCAGAGTATTTGGGACGACCGA	1024
Dd	1243	AGAAGACGAGCGTGTGATAATAGTGGTGGAGGATACCCGAGTACTGGGATGACGAGA	1302
Qy	1025	AGAAGACGACTATTACTGCAATACGCGCGCAGATGTCCGCGCTACTTACGGCTTAG	1084
Dd	1303	AGAAGGCGAGCTCTAGAGCGCTACGCGCTCGACGATGTGAGGCTTACCTACGCGCTCG	1362
Qy	1085	CCGAGAGATATTCCCGTTTCTATCCAGTTTCTGCTAGCTAAACAGGTCTCCCACTAGACC	1144
Dd	1363	CGGAAAGATGCTACCGTTTCCCAATACAGCTCTCCACTGTATCGGGTGTGCTCTCGACC	1422
Qy	1145	AGGTAGGTGCGATGAGTTTGGCTTTCGACTTGAATGGTACCTGATACGCGCGCGGTTTA	1204
Dd	1423	AGGTAGGTGTATGGCGTAGGCTTCCGCTTAGAGTGGTATCTCATGCGTGCAGCTTACG	1482
Qy	1205	AGATGAAGAGCTTGTGCGGAACCGGTTGAGCGCCCAAGAGACTTACGTGGCGCTA	1264
Dd	1483	ATATGAACGAGCTGTGTGCGGAACCGGGTGGAGAGAGGGGGGAGAGCTACAAGGGTGCAG	1542
Qy	1265	TAGTTCTTGGCCGTGTGAGGCGGTGACAGAGATATAGCCGTACTGCTAGCTTTAGCTCGA	1324
Dd	1543	TAGTTTAAAGCCCTCTAAGGGAGTCCATGAGAAATGTTGTGGTCTCGATTTAGTTCCA	1602
Qy	1325	TGTACCCAAACATCATATGAATAAGTACAAATGTTGGTCTCTGACACGCTTGTGAGGCGCTGTG	1384
Dd	1603	TGTACCCGAGCATATGATGAATGATACACGTCGTGGCGCCCGGACACTATAGTCCAGACCCCT	1662
Qy	1385	AAAGCTGTGGCGAGT---GTGTTGCTGGAGGCGCCCGAGGTCAAGCACAGGTTCCGTA	1441
Dd	1663	CGAGTGGCCAAAGTACGCGCGCTCTATGTAGCCCCCGAGGTGGGACCGGTTCCGTC	1722
Qy	1442	GGTGTCCGCGCCCGCTTCTTCAAGACAGTCTTTGAGNGGCTGTAGAGCTTCTGTAAGCGTG	1501
Dd	1723	GCTCCCGCGCGAGCTTCTTCAAGACCGTCTCGAAGCCTACTGAGCTACGCCACAGG	1782
Qy	1502	TCGCTGTGAANTAAAGAAAGTATCTCCGATAGCCAGATATCGACTGTGGATGAAA	1561
Dd	1783	TAAAGAGAAAGATGAAGAGTTTCCGCTCGACAGCCCGAGTACAGGCTCTACGATGAGC	1842
Qy	1562	GCAGAGCGGTTGAAGTCTTTCAAAACGCTAGTTACGGCTACATGCTGGGTTGAGCGGCG	1621
Dd	1843	GCAGAGCGCTCAAGGTTCTTGGAAACGCGAGCTATGGCTACATGGGGTGGAGCCATG	1902
Qy	1622	CTAGGTGGTATTGTCAGGAGGTGCGCAAGGCTGTCAACGCTTGGGCTAGGCAACCTCATAC	1681
Dd	1903	CCGCTGTACTGCAACCGCTGCGCGAGGCTGTCAAGCTTGGGCGGCTAACCTTATAC	1962
Qy	1682	GCACCGCATCAACATAGCTCGTAAACTAGGCTCAAGGCTCAAGGTGATCTACGGTGACAGATT	1741
Dd	1963	TGACAGCTATCGAGTATGCGAGGAGCTCGGCTTAAAGGTTATATATGAGAGACCGACT	2022
Qy	1742	CGCTCTTGTGACTTATGATCCGAGAGGTTGGAATTTTCATCAAAATTTATAAGGAGG	1801
Dd	2023	CCCTCTTGTGCTTATGACAAAGGAGGTTGGAAGCTGTAGAGTTTGTTCGAGNAGG	2082
Qy	1802	AGCTGGGTTCGAATCAAGCTAGAGAGGTTGTACAAACGCTTATTTTACAGAGGCTA	1861
Dd	2083	AGCTGGCTTTGAGATTAAGATAGACAGATCTACAGAAAGTGTCTTTCAGGAGGCTA	2142
Qy	1862	AGAAGAGTACGCTGGCTTCTCGAGGACGAGCTATAGATATGTCGGTTCGAGGCTG	1921
Dd	2143	AGAAGCGCTATGTAGTCTCTCGAGGACGAGCTATAGACATCGTGGCTTTGAAGCAG	2202
Qy	1922	TACGTGGCATTTGTTGTAACCTCGCCAGAGGTTTCAGACTAAGCTTGTTCGAAATAGTAT	1981
Dd	2203	TCGCGCGCACTGGTTCGAGCTGGCTAAGAGGTTGAGAGAGGCGGCTGAGATAGTGT	2262
Qy	1982	TGAAGACGAGTGAAGGCTGTAGAGTACGCTCAGGAAGATTGTGGAAGAGTTGG	2041
Dd	2263	TGAATACGCGGAACGTTGGACAGGCTATAAGCTACATAAGGAGGTAATAAGCAGCTCC	2322
Qy	2042	AGAGGCGAAGGTTCCCATAGAGAGCTTGTAACTCTGGAAGCCCTTAGTAAGCGTCTTG	2101
Dd	2323	GCAGGCGAAGGTTGCCAATAACAAAGCTTATCATATGGAAGACGCTGAGCAAGAGATAG	2382
Qy	2102	AGAGTACACAACGAGGACCAACACGCTGTTGACGAGGAGGATGCTGTCCAGCAGGCT	2161
Dd	2383	AGAGTACAGCATGACGCGCTCTGTGATGGCTGACGCGCTATGAAGGAGGCGAGGCT	2442
Qy	2162	ACCGGTAAGCCGACGCAAGATAGGATATGTAATAGTAAGGAGGTTGTGGCCGCTATCA	2221
Dd	2443	ACGAGGTGTCTCCCGCGATAGGTGGGCTACGTATAGTTAAGGTAAGGAGTGTGT	2502
Qy	2222	GTCAAAGAGCATGGCCCATCTTCTATGTCGTAAGATCTCTAGCCAGATAGACGTGACCTACT	2281
Dd	2503	CCAGCAGGCGCTACCCCTACTTCTCATGGT---TCATCCATCGACCATCGACGCTCAACT	2559
Qy	2282	ATGTTGACCAACAAATCATCCCGCTGCAATTCGAGATATCTGGGCTACTTTGGCATCACCG	2341
Dd	2560	ATATTGACCCAGATAGTGGCGGCTGCTCTGAGGATCTCTTCTACTTCTGAGGTACCG	2619
Qy	2342	AGAAGAGCTGAAAGCAAGTGAACCTGGGCGAGAGACTCTCTTTCGACTTTCTAGGCCAAGA	2401
Dd	2620	AGAACAGCTCAGGCGGCGGTACGCTGACAGAGAGGCTCTTCTGACTTCTTCGCTCAA	2679
Qy	2402	AGAGCAAG	2409
Dd	2680	AGAAATAG	2687

RESULT 7

PYRDNAPB 3069 bp DNA linear BCT 08-FEB-1999
 LOCUS Pyrodicticum occultum polB gene for DNA polymerase II (family B)
 DEFINITION (alpha-like DNA polymerase), complete cds.
 ACCESSION D38574
 VERSION D38574.1 GI:807829
 KEYWORDS polB; DNA polymerase II; alpha-like DNA polymerase; DNA polymerase family B.
 SOURCE
 ORGANISM Pyrodicticum occultum
 Pyrodicticum occultum
 Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
 Pyrodictiaceae; Pyrodictium.
 1 (bases 1 to 3069)
 Uemori, T., Ishino, Y., Doi, H. and Kato, I.
 The hyperthermophilic archaeon Pyrodicticum occultum has two
 alpha-like DNA polymerases
 J. Bacteriol. 177 (8), 2164-2177 (1995)
 95238290
 7721707
 REFERENCE
 MEDLINE 2 (bases 1 to 3069)
 PUBMED Ishino, Y.
 AUTHORS Direct Submission
 JOURNAL Submitted (21-OCT-1994) Yoshizumi Ishino, Takara Shuzo Co., Ltd.;
 Title Seto 3-4-1, Otsu, Shiga 520-21, Japan (Tel:0775-43-7215,
 Fax:0775-43-2494)
 FEATURES
 Location/Qualifiers
 1..3069
 /organism="Pyrodicticum occultum"
 /mol_type="genomic DNA"
 /strain="DSM2709"
 /db_xref="taxon:2309"
 /clone="pFO500-II"
 276..2687
 /gene="polB"
 276..2687
 /gene="polB"
 /EC_number="2.7.7.7"
 /codon_start=1
 /transl_table=11
 /product="DNA polymerase"
 /protein_id="BAA07580.1"
 /db_xref="GI:807830"
 /translation="MTETIEFVLDSSYEILGKEPVVILNGITLDRKRVLLDHRFRP
 YFYLARGEDWVEEPAASIRLSVYKSPILIDAKPLDKRIFRGPRAKVIITIMIPES
 VHRVREAKLEGWEDLEADIEFARMYLDKELYFTVIRIPVEDAGRNPGFRVDR
 YKAGDPEPLADITRIIDLPMLRVAFDIEVYSRSGSPNPAFDVITVLSRDSGKERL
 IEAGHDRLVRFVEYVRAFPDIIIVGNSNHFDPYLMERARLGIKLDVTRVG
 AEPTTSVGHVSQGLNVDLYDAEMPEIKMTELEVAEYLVGMKKSERVILEWRV
 IPEYMDKXKQLLEVALDDVRATYGLAEKMLPEALQISTVTGVLDOVGAMGTGR
 LEWYLMRAADMELVNVRERGESYKGAIVLKLPLGVHENVVLDLDFSMYPSMIK
 INVGDITVDPSCEPKYGCYVAPEVGHFRPSPPGFFKTVLENLLKURRVKRMK
 EFPDSEYRLYDRQALVLANASYGMGSHAWYCKRCREAVTAWGNRLIITAI
 EYARLGLKVIYDSDLSLFVYDKVEKLEIFVEKELAFEIKIDIKYKVFTEAKK
 RYVGLLEDGRDIDVGFVAVRGDWCLEAKVEQAAEIVLNTGNVDXIAISYIREVAKL
 REGKVPITKLIWTKLSRIEVEYHDAPHVMAARMKEAGYEVSPGDKVGYVIVKSG
 SVSSRAFYEFWVDSTIDVNYIDHQLVPAALRLSYFYTEKQLKAAATVQBSLEDF
 FASKK"
 730 a 746 c 980 g 613 t

source

Query Match 44.4%; Score 1071.2; DB 1; Length 3069;
 Best Local Similarity 66.5%; Pred. No. 2.5e-286;
 Matches 1602; Conservative 0; Mismatches 788; Indels 18; Gaps 4;
 11 TTGTAATCAGGTTTACCTCTAGCTACGAGGTTGTGGTAAAGACCTCAGGTAATCA 70
 289 TAGAGTTCGTCTGCTAGACTCTAGCTACGAGTACTGGGGAAGAGCCCGTATGATCC 348
 71 TATGGGATATGCTGAGAACCGGAGAGGCTAGCTCATGACAGCTCTTTCCGCCAT 130
 349 TCTGGGATTAACGCTTACGCTAAACGCTGCTGCTCTAGACCAACCGCTTCGCCCT 408

gene

cds

131 ACTTCTATGCGCTGTGTCACCGCGCGCGATCCTA-----AGCAGGTAGCACAACTGA 184
 409 ACTTCTACGCGCTCATAGCGCGCGGCTATGAGGATATGCTGAGGAGATAGCAGCTTCCA 468
 185 TTCTGTCNTGTAGTAGCCAAAGAGCCCGATATAGGTGTAGAGGATGACAAAGAGGAGT 244
 469 TAAGGAGGCTTAGTGTGTCTCAGAGTCCGATATAGATGCCAAGGCTCTTGTAAGAGGT 528
 245 ACTTCGGGAGGCTCGCTAGGCTTCTACGATATTCGACCGCTGTACCCGAGGCTGTAGG 304
 529 ACTTCGGCAGCGCCCGTAAGCGGTGAAGATATACCACTATGATACCCGAGTCTGTAGAC 588
 305 AGTATCCGCAACTCGTAAAGACGTTGATGTTGCTGAGGATGTTCTAGAGGCGGATATAC 364
 589 ACTACCGCGAGCGGTGAAGAGATAGAGGTTGTGAGGACTCTCTCGAGGCGAGATATTA 648
 365 GCTTCGCTATGCGCTATCTCATAGATCACGATCTATTTCTTTCACCTTGTTACCGGTAG 424
 649 GGTTCGATAGATATCTGATAGATAAGAGGCTCTACCCGTTTACCGGTTTACCGGATCC 708
 425 AGGCTGAGCCCTCGAGAACAGATCGGCTTCGCTGTCGACAGGATATACCTTCTTGAAGA 484
 709 CCGTAGAGGATCGGCGCGCATTCAGGCTTCGCTGTCGCTGTCACAGGTTCTGCTG 768
 485 GCAGGCGGAGCCACTTTATGCTGAGGCTCTCGCACCAACCAAGCTTCCCGATCTTAGGA 544
 769 GCGACCCGAGGACCCCT-----AGCGGATATAGCGGATCGACCTTCCCGGATGAGCG 822
 545 TACTCGGCTTCGATATTAAGTTTATAGCAAGCAAGGTCGCGCTCCAGAGCGCGATC 604
 823 TGGTAGCTTTTGATATAGAGGTGTATAGCAGGAGGCGGAGCCCTAAACCTTCAAGGAGT 882
 605 CTGTAATAGTATAGTCTGTGAAGACTGACGATGCGCATGAGGTGCTTATTCATTGTCAGAG 664
 883 CAGTGATAATAGTCTGCTGAGGACACGCGGCGGACAGGAGGCTCATAGAGCTGAAG 942
 665 GCAAGACGATCGAAGAACCGATACGCGAGTTTGTAGAGTACGTGAAGAGGTATGACCCG 724
 943 GCCATGACGACAGGAGGTTCTGAGGAGGTTCTGATAGTACGCTGAGAGGCTTCGACCCG 1002
 725 ACATAATAGTCGGTTATAACAACATCATTTCCATTGCGCTTATCTTTTGGCGCGCGCC 784
 1003 ACATAATAGTGGCTATTAACAGTAACCACTTCGACTGGCCCTACCTAATAGGAGCGGCC 1062
 785 GCATCTTAGCATAAAGCTTGTAGTGAAGAGTTGCGCGCGAGCCACCACTAGGCG 844
 1063 GTAGGCTCGGATTAAGCTCGAGCTTACACGCGCTGTGGGGGAGAGCCACCAACGAGCG 1122
 845 TACATGGGCGCTCTGCTCCCTGCGAGGCTTAAACGTAGATCTGTACGACTATGCGGAG 904
 1123 TCTAGCGCCAGCTCTCGCTGCGAGGTAGGCTGAACGTGACCTCTACGACTATGCGGAG 1182
 905 AGATGCCAGAGATCAAGATAAAGAGTCTCGAGAGGTCGACAGTATCTAGCGCTGATGA 964
 1183 AGATGCCAGAGATAAAGATGAAGACGCTTGAAGAGGTAGCGGAGTACCTTAGGCGTTATGA 1242
 965 AGAGAGTGAACGCGTTATCATCAATGCTGGAGAGTTCAGACTATTTGGGAGCGCCGA 1024
 1243 AGAAGAGCGAGCGGTGATTAATAGATGTTGGAGGATACCCGAGTACTGGGATGCGAGA 1302
 1025 AGAAGAGACCACTATTACTGCAATACGCGCGGACCATGCTCCGCGCTACTTACGGCTTAG 1084
 1303 AGAAGAGCGAGCTGTAGACGCGCTACGCGCTGACGATGTGAGGGGCTACTTAGGCGCTCG 1362
 1085 CCGAAGAGATATTCGCGCTTGTGCTATCCAGTTGTGCTAGCTACAGGTCTCCCACTAGACC 1144
 1363 CGGAAAGATGCTACCGTTCCGCTACAGCTCTCCACTGTTAGGGGTGTGCTCTCGAC 1422
 1145 AGGTAGTGGGATGAGTGTGGCTTTCGACTTGAAGTGTACCTGATACCGCGGCGGTTTA 1204
 1423 AGGTAGTGTCTATGGGCGTAGGCTTCGCGCTAGGTGTTATCTCATGCGGTGAGCGCTAG 1482
 1205 AGATGAAGAGCTTGTGTCGCAACCGCGCTTGTAGCGCCGACGAGAGACTTACCGTGGCGCTA 1264

```
1483 ATATGACGAGCTGGTCCGACACCGGTGGAGAGGAGGGGGAGAGCTACAAGGGTGCAG 1542
1265 TAGTTCTTGAGCGCTGAGAGGGGTGACAGAGATATAGCCGCTACGCTTATAGTCGA 1324
1543 TAGTGTAAAGCCCTCAGAGGAGTCCATGAGAAAGTTGGTGGCTCGATTCAGTTCCA 1602
1325 TGTACCCAAACATCATGATAAAGTACAAATGTTGGTCTGTGACACGCTTGTGAGCCCTGGTG 1384
1603 TGTACCCGAGCATATGATAAAGTACAAAGTGGGGCCCGACACTATAGTCGACGACCCCT 1662
1385 AAAAGTGTGGGAGT---GTGGTGTGGGAGGCCCCGAGGTCACGACAGTTCGGTA 1441
1663 CGGAGTCCCAAGTACGGCGGCTGCTATGAGCCCCCGAGGTCGGGACCGGTTCCGTC 1722
1442 GGTGTCCGCGCGCTCTTCAAGACAGTCTTCAAGAGGCTGTAGAGCTTCGTAAACGTC 1501
1723 GCTCCCGCCAGGCTTCTCAGAACCGTGTCTGAGAACCTACTGAAAGCTAGCGGACAGG 1782
1502 TGGGTGTGAATGAAGAGTATCTCCGGATAGCCCGAGATATGCACTGTTCGGTGA 1561
1783 TAAAGGAGAGATGAAGAGTTCCTCCGCTGACAGCCCCGAGTACGCTCTACGATGAGC 1842
1562 GGCAGAGGCGTCTGAAGTCTTCCAAACGCTAGTACGCTACATGGGTTGAGCGGCG 1621
1843 CCCAGAGGCGCTCAGGCTTCTCCGAAACCGGAGCTATGGCTACATGGGTTGAGCCATG 1902
1622 CTAGGTGTATTTGAGGAGGTGCGCAAGGCTGTACGCGCTTGGGTGAGCACCTCATAC 1681
1903 CCGCTGTGTCTGCAAAAGCTGCGCGGAGCTGTACAGAGCTGGGCGGTAACTTATAC 1962
1682 GCACCGCATCAATAGCTGTAACTAGAGCTCAGGTGATCTAGCGTGACACAGATT 1741
1963 TCACAGCTATCGAGTATGCCAGGAGTTCGCGCTTAAAGTTATATATGAGACACCGACT 2022
1742 CGCTTCTGCTGACCTATGATCCGAGAGGTGGAAATTTCAACAAATATATAAGGAG 1801
2023 CCTCTTCTGCTGTCTATGACAGAGGAGAGGTTGAGAGCTGTAGAGTTTGTGAGAGG 2082
1802 AGCTGGGCTTCAAAATCAAGCTAGAGAGGTGTACAAACGCTTATCTTTACAGAGGCTA 1861
2083 AGCTGGGCTTGTAGATAAGATAGACAGATCTACAAAGAAAGTGTCTTACCGAGGCTA 2142
1862 AGAGAGTACGCTGGGCTCTTCGAGGACGAGCTATAGATATTTGTCGGTTTCAGGCTG 1921
2143 AGAGCGCTATGTAGTCTCTCGAGGACGAGCTATAGACATCTGTCGGCTTTGAGAGCAG 2202
1922 TACGTGGCGATTGGTGAATTCGCCAAGGAGGTTTCAGACTAAGGTTGTGAAATAGTAT 1981
2203 TCCGCGGCGACTGGTGGAGCTGGCTAAGGAGGTGCAGGAGAGGCGGCTGAGATAGTGT 2262
1982 TGAGAGCAGTGTGAGTGAACAGGCTGTAGAGTACGTCAGAGATTTGAAAGAGTTGG 2041
2263 TGAATACGGGAACTGGACAGGCTATAAGCTACATAAGGAGGTAATAAGACAGCTCC 2322
2042 AGGAGGCAAGCTTCCCATAGAGAAGCTTGTAACTCGAAGACCCCTTAGTAAGCGCTTG 2101
2323 GCGAGGCAAGGTGCAATACAAAGCTTATCATATGGAAGACGCTGAGCAGAGGATAG 2382
2102 AGGAGTACAAACGAGGACACACAGCTGTGTCAGGAGAGAGATGCTGTACAGAGGCT 2161
2383 AGGAGTACAGCATGACGCGCTCATGTGATGGCTGCACGCGCTATGAGAGGAGGAGCT 2442
2162 ACCGGGTAAAGCCAGGCGACAGATAGGATGTGTAAGTGAAGGTTGGGCGCTATCA 2221
2443 ACGAGGTGTCTCCCGCGCATAGGTGGCTACGTATAGTTAAGGTTAGCGGAGTGTGT 2502
2222 GTCAAGAGCATGGCCATCTTCTAGTCAAGGATCTCTAGCCAGATGAGCTGACCTACT 2281
2503 CCAGCAGGCGCTTACCCCTACTTCTATGTT---TGATCCATCGACCATCGAGCTCACTACT 2559
2282 ATGTTGACCAACCAATCATCCGCGCTGCAITTGAGAACTCTGGGCTACTTTGGCATCCG 2341
```

Db 2560 ATATTGACCAACAGATAGTCCCGCTCTCTGAGGATACCTCTCTACTTTGAGTACCG 2619

Qy 2342 AGAAGAGCTCAAGCAAGTGCACACTGGGAGAGAGTCTCTCTCGACTTTCTTAGCCACGA 2401

Db 2620 AGAAGAGCTCAAGCGCGGCTACGCTGCGAGAGAGCTCTCTCGACTTTCTTAGCCCTCA 2679

Qy 2402 AGAGCAAG 2409

Db 2680 AGAATAG 2687

RESULT 8

LOCUS 117571 2430 bp DNA linear PAT 07-OCT-1996

DEFINITION Sequence 3 from patent US 5491086.

ACCESSION 117571

VERSION 117571.1 GI:1597926

KEYWORDS

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 2430)

AUTHORS Gelfand, D.H. and Wang, A.M.

TITLE Purified thermostable nucleic acid polymerase and DNA coding sequences from pyrodicticum species

JOURNAL Patent: US 5491086-A 3 13-FEB-1996;

FEATURES

source 1. 2430

Location/Qualifiers

BASE COUNT 604 a 579 c 762 g 485 t

ORIGIN

Query Match 44.3%; Score 1069.6; DB 6; Length 2430;

Best Local Similarity 66.5%; Pred. No. 6.7e-286;

Matches 1601; Conservative 0; Mismatches 789; Indels 18; Gaps 4;

Qy 11 TTGATTACACGGTTTATAGACTCTAGCTACGAGGTGTGTGTAAGAGCCCTCAGGTATCA 70

Db 14 TAGAGTTGCTGCTCTAGACTCTAGCTACGAGATCTGCGGAGAGGAGCGGTAGTATCC 73

Qy 71 TATGGGTATTTGTGAGAACGGCGAGAGGTAGTCTCTATTGACAGGTCTTTTCGCCCAT 130

Db 74 TCTGGGGATTAACCTTTGACGGTAAACGCTGTCTGCTCTCTAGACCACCGCTTCGCGCCCT 133

Qy 131 ACTTCTAGCTGCTCTTGTGACCCGCGCGCCGATCCTA-----AGCAGGTAGCACACGTA 184

Db 134 ACTTCTAGCCCTCATAGCCCGGCGCTATGAGATATGTTGGAGGATAGCAGCTTCCA 193

Qy 185 TTCGTGCAATTGAGTAGGACCAAGAGCCCGGCTATAGGTGTAGAGGATGACAGAGGAAGT 244

Db 194 TAAGAGGCTTAGTGTGTCTCAAGAGTCCGATATAGATGCAAGCTCTTTGATAAGAGGT 253

Qy 245 ACTTCGGAGGCTCTGAGGTCTTAGGTATTCGACCGTGTACCGAGGCTTTAGGG 304

Db 254 ACTTCGGAGGCTCTGAGGTCTTAGGTATTCGACCGTGTACCGAGGCTTTAGGAC 313

Qy 305 AGTATCGCAACTCGTAAAGAACGTTGATGTGTGAGCATGTTCTAGAGCGCGATATAC 364

Db 314 ACTACCGCGCGGCTGAGAGATAGAGGTTGAGGACTCCCTCGAGGATATATA 373

Qy 365 GCTTCGTATCGGTATCTCATAGATCAGATCTATTTCTTTTCACTGAGTCCGCTAG 424

Db 374 GGTTCGAATGAGATATCTGATAGTAAGAGGCTTACCGGTTACCGGTTACCGGATCC 433

Qy 425 AGGCTGCGCCCTCGAGAACAGAGTGGCTTCGCTGTCGACAGGATATACCTGTTAAGA 484

Db 434 CCGTAGAGATGCGGGCGCGATCCAGGCTTCGTTGTTGACCGGTGTCACAGGTTGCTG 493

Qy 485 GCAGCGCGGACCCACTTTATGTTGAGGCTCTCGCACCAACCAAGCTTCCCGATCTTAGGA 544

Db 494 GCGACCGGAGCCCT-----AGCGGATATAACGGGATCGACCTTCCCGCATGAGGC 547

Qy 545 TACTCGGCTTCGATATTGAAGTTTATAGCAAGCAAGGCTCGCGGCTCCAGAGCGCGATC 604

Db	548	TGCTAGCTTTTGTATAGAGGTGTATAGCAGGAGGGAGCCCTAAACCTTCGCAAGGGATC	607
QY	605	CTGTATATAGTATAGTGTGAAGACTGAACGATGGCGATGAGGTCTATTATTCATTTGCAGAGG	664
Db	608	CAGTGTATATAGTGTGCTGAGGACACAGCGAGGCAAGGAGAGGCTCATAGAGCTGAAG	667
QY	665	GCAAGACGATCGAAGAACCATACCGGAGTTTGTAGAGTACGTGAGAGGTATGACCCCG	724
Db	668	GCCATGACGACAGGAGGTTCTGAGGAGTTCGTAGAGTACGTGAGAGGCTTCGACCCCG	727
QY	725	ACATATAGTTCGGTGTATTAACAACAATCATTTCCATTGGCTTATCTTTTGGAGCGCGCC	784
Db	728	ACATATAGTTCGGCTATTAACAGTAAACACTTCACCTGGCCCTACCTAATGGAGCGCGCC	787
QY	785	GCATCTTAGGCTAAGAGCTTGTATGATGACCTAGAAGATTGGCGCGGACCCACCATAGCG	844
Db	788	GTAGGCTCGGGATTAACCTCGACGTTACACCGCGTGTGGGGCAGAGCCACCCACAGG	847
QY	845	TACATGGGACGCTCTGTCCCTGGCAGGCTTAACGTAGATCTGTACGACTATGCGGAG	904
Db	848	TCATCGGCGACGCTCTCGGTGACGGTAGGCTGAACGTGGAACCTCTACGACTATGCGGAG	907
QY	905	AGATGCCAGAGATCAAGATAAAGAGTCTCGAGAGGTTCGAGAGTATCTAGCGCTGATCA	964
Db	908	AGATGCCGAGATAAAGATGAAGACGCTTGAGAGGTACGAGTACCTAGGCGTTATGA	967
QY	965	AGAAGAGTGAACCGCTTATCATCAATGTGTGGAGATTCCAGACTATTTGGGACGACCGA	1024
Db	968	AGAAGAGCGAGCGTGTGATAATAGAGTGTGTGGAGGATACCCGAGTACTGGGATGACGAG	1027
QY	1025	AGAGAGACACCATTAATCTGCAATAGACCGCGGACGATGTCCGCGTACTTACGCGTTAG	1084
Db	1028	AGAAGAGCGAGCTGTAGAGCGCTACGCGCTCGACATGTGAGGGCTACTCTACGCGCTCG	1087
QY	1085	CCGAGAAGATPATTCGCTTGTCTATCCAGTTGTGCTGACGTAAACAGTCTCCCACTAGACC	1144
Db	1088	CGAAGAATGTCTACGTTTCGCTACGCTACGCTCTCCTGTTACGGGTGTGCTCTCGACC	1147
QY	1145	AGTAGGTGGATGAGTGTGGCTTTTGCATTTGAATGATGATGATGATGATGATGATGATGAT	1204
Db	1148	AGTAGGTGTCTATGGCGTAGGCTTCCGCTAGAGTGATCTCTATGCTGTGAGCTAGC	1207
QY	1205	AGATGAAGAGCTTGTGCGGACCGGTTGAGCGCCGAGAGACTTACGCTGAGGCTA	1264
Db	1208	ATATGACGAGCTGTGTCGGAACCGGTTGAGAGAGGGGGAGAGCTACAAGGGTGCAG	1267
QY	1265	TAGTCTTGTAGCGCTTGTAGAGCGTGCACAGAAATATAGCCGTACTCGACTTTAGCTGA	1324
Db	1268	TAGTGTAAAGCCTCTCAAGGAGTCCATGAGNATGTTGTGTGCTCGATTTCACTTCCA	1327
QY	1325	TGTACCCAAACATCATGATAAAGTACAATGTTGGTCTGACACGCTTGTAGGCGCTGGTG	1384
Db	1328	TGTACCCGAGCATATGATAAAGTACAACGTTGGGCGCCGACACTATAGTCGACGACCCCT	1387
QY	1385	AAAAGTGTGCGGAGT---GTGGTGTGTTGGAGCGCCCGGAGCTCAAGCAGGTTCCGTA	1441
Db	1388	CGGAGTGTCCCAAGTACGCGCGCTGTATGTAGCCCGGAGTGTGCGGACCGGTTCCGTC	1447
QY	1442	GGTGTCCCGCGGCTTCTTCAAGACAGATTCTTGAAGAGGCTGTAGAGCTTCTGTAAGCGTG	1501
Db	1448	GCTCCCGCGCAGGCTTCTTCAAGACGCTGCTCGAGAACCTACTGAGTACGCGCCACAGG	1507
QY	1502	TGGGTGCTGAATGAAGAGTATCTCCGATAGCCGAGATPATCGACTGTGTGATGAAA	1561
Db	1508	TAAAGGAGAAGATGAAGAGTTCGCGCTGACAGCGCCGAGTACAGGCTCTACGATGAGC	1567
QY	1562	GCAGAGGCGCTTGAAGTCTTTCGAAAGCTTGTAGAGTACGCTTACGCTGAGTGGTGGAGCGG	1621
Db	1568	GCAGAGGCGCTTGAAGTCTTTCGAAAGCTTGTAGAGTACGCTTACGCTGAGTGGTGGAGCGATG	1627
QY	1622	CTAGGTGTGTTTGCAGGAGTGCACAAAGGCTGTACGCTTGGGGTAGGCACTCATAC	1681
Db	1628	CCCGCTGTGTACTGCAAAACGCTGCGCCGAGGCTGTACAGCCCTGGGGCGGTAACTTATAC	1687
QY	1682	GCACCGCCATCAACATAGCTCGTAAACTAGCCCTCAGGCTGATCTACGCTGACACAGATT	1741
Db	1688	TGACAGCTATCGAGTATCGCAGGAGCTCGGCTTAAGGTTATATATGAGACACCGACT	1747
QY	1742	CGCTTTCTGTACCTTATGATCCGAGAAAGGTGGGAAAATTTTCATCAAAATTAATAAGAGG	1801
Db	1748	CCCTTTCTGTGCTTATGCAAGAGGAGAGGTTGAGAGCTGATAGAGTTTCTCGAAGG	1807
QY	1802	AGCTGGGCTTCGMAATCAAGCTAGAGGTGTTTAAACCGCTTATCTTTCAGAGGCTA	1861
Db	1808	AGCTGGGCTTTGAGATAAAGATAGACAGATCTCAAGAAAGTGTCTTTCAGGAGGCTA	1867
QY	1862	AGAAGAGTACGCTGCGCTTCTCGAGGACGACGCTATAGATATTTGTCTGCTTCAGGCTG	1921
Db	1868	AGAGCGCTATGTAGTCTCTCTCGAGGACGAGCTATAGACATCTGTGGCTTTGAGGAG	1927
QY	1922	TACGTGGCGATTGTGTGAATCCCAAGGAGGTTTCAGACTAAGGTTTGTGAAATAGTAT	1981
Db	1928	TCGCGCGGCTGTGTGCGAGCTGCTAAGGAGGTGCAGGAGAGGCGCTGAGATAGTGT	1987
QY	1982	TGAGACGAGTGTGTGAACCAAGGCTGTAGAGTACGCTCAGGAGATTTGTGAAAGATTGG	2041
Db	1988	TGAATACGGGAGACGTGGACCAAGCTATAAGCTACATAAGGAGGTTAATAAGCAGCTCC	2047
QY	2042	AGGAGGCAAGGTTCCCATAGAGAAGCTTTGTAATCTCGAAGACCTTTAGTAAGCGTCTTG	2101
Db	2048	CGGAGGCAAGGTTGCCAATTAACAAGCTTATCATATGGAAGACCTGAGCAAGGATAG	2107
QY	2102	AGGAGTACAAACGAGGACCAACACGCTGCTGTGAGCGAAGAGGATGCTGTACAGGCT	2161
Db	2108	AGGAGTACGACATGACGCGCTCATGTGTGAGGCTGACGCGGTATGAAGGAGGACGCT	2167
QY	2162	ACCGGTAAGCCGAGGCGCAAGATAGGATGTGTAATAGTGAAGGTTGGTGGCGGTATCA	2221
Db	2168	ACGAGGTGTCTCCCGGCGATAGGTGGGCTACGCTAGTAAAGGTTAGCGGAGTGTGT	2227
QY	2222	GTCAAGAGCATGCGCCATCTTCAATGCTCAAGGATCTTAGCCAGATAGAGTACCTACT	2281
Db	2228	CCAGCAGGCTTACCCCTACTTCTCATGCT---TGATCATCGACCTCGACCTCACTACT	2284
QY	2282	ATGTTGACCCAAATCATCTCCGCTGCTGATGAGAACTCTGGGCTACTTTTGGCATCCG	2341
Db	2285	ATATTGACCCAGATAGTCCGCTGCTCTGAGGACTACTCTCTACTTCCGAGTCCAG	2344
QY	2342	AGAAGAGCTGAAAGCAAGTGCACCTGGCAGAGAGCTCTCTTTCGAGCTTTCTAGCCAGA	2401
Db	2345	AGAAACAGCTCAGCGCGGCTACGGTGCAGAGAGCCCTCTTCTGAGCTTTCTTCCCTCAA	2404
QY	2402	AGAGCAAG 2409	
Db	2405	AGAAATAG 2412	
RESULT 9			
LOCUS	117570	2430 bp	DNA linear PAT 07-OCT-1996
DEFINITION	Sequence 1 from patent US 5491086.		
ACCESSION	117570		
VERSION	117570.1	GI:1597925	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 2430)		
AUTHORS	Gelfand, D.H. and Wang, A.M.		
TITLE	Purified thermostable nucleic acid polymerase and DNA coding sequences from pyrodicticum species		
JOURNAL	Patent: US 5491086-A 1 13-FEB-1996;		
FEATURES	Location/Qualifiers		
source	1..2430		
	/organism="unknown"		

BASE COUNT	747 a	541 c	657 g	485 t	
ORIGIN					
Query Match	43.3%	Score 1045.6;	DB 6;	Length 2430;	
Best Local Similarity	65.9%;	Pred. No. 3.2e-279;			
Matches 1586;	Conservative	0; Mismatches 804;	Indels 18;	Gaps 4;	
QY	11	TTGTATTCACGGTTTATAGACTCTAGCTACGAGGTGTGTTGGTAAAGAGCCCTCAGGTAATCA	70		
DB	14	TAGATTCTGTGCTTCTTGTATTCAGCTACGAGTTGTAGGAAAGGCCGCGTAATCATAC	73		
QY	71	TATGGGGTATTGCTGAGAAACCGGAGAGGGTAGTCTCTCATTTGACAGGTCTTTTCGCCCAT	130		
DB	74	TATGGGGTGTAAAGCTAGACGGTAAACGATAGTCTCTCTTGTATAGAGGTTTATAGGCCCT	133		
QY	131	ACTTCTATCGCTGTGCTTGCACCGGGCGCCGATCCTA-----AGCAGGTAGCACACGTA	184		
DB	134	ACTTCTATGCACTCATATCCCGGACTACGAGGTAAAGCCGAGGAGGTAGTAGCTGTCTA	193		
QY	185	TTCTGTGATGATGAGTACGACCAAGAGCCCGATTTATAGGTGTAGAGGATGACAGAGGAGT	244		
DB	194	TTAGAGGGCTTAAGTATGGCAAGAGCCCAATATAGAGCAAGGTGGTTAGTAAGAGT	253		
QY	245	ACTTCGGAGGCTCTGAGGGTCTTACGTATTCCGACCGGTCTACCCGAGGCTGTTTAGGG	304		
DB	254	ACTTCGGAGGCTCTGAGGGTCTTACGTATTCCGACCGGTCTACCCGAGGCTGTTTAGGG	313		
QY	305	AGTATCGCGAACTCGTAAAGAAAGCTTGTATGGTGTGAGGATGTTCTAGAGGCGGATATAC	364		
DB	314	AATATAGAGAGGCTGTAAGAAAGCTGGAAGCGGTGGAAGACTCTCTAGAGCAGACATAA	373		
QY	365	GCTTCGCTATGCGCTATCTCATAGATCAAGATCTATTTCCTTCACTGCTACCGGTAG	424		
DB	374	GGTTCGGGATGAGGATCTTAATCGAAGAGCTCTACCGTTTCAGACATACCGTGTCA	433		
QY	425	AGGCTGAGCGCCCTCGAAGCAAGATGGGCTTCGGTGTGACAAAGGTATACCTGGTTAAGA	484		
DB	434	GAGCGAGAGACGCTGGAGCGAGCCCTGGTTTCGGTGTAGACTCGGTATAC-----ACTA	487		
QY	485	GCAGCGCGAGCCACTTTATGTTGAGGCTCTCGACCAACCAAGCTTCCGATCTTAGGA	544		
DB	488	TAGTTGAGAGCCACAGAGCTATTGCGGACATACATAGATATAGATATACAGAGATCGGTG	547		
QY	545	TACTCGCGTTCGATATTGAGTTTATAGCAAGCAAGGCTCGCGGTCCAGAGCGGATC	604		
DB	548	TGCTCGGTTTCGACATAGAGTCTACAGTAAGAGAGGAGCCCTAACCCGCTCCGCGACC	607		
QY	605	CTGTAAATAGTATAGCTGTGAAGACTGACGATGCGGATGAGTGTCTATTTCATTTCCAGAGG	664		
DB	608	CGGTCAATAATCTCGATTAAGGACAGCAAGGCGGACGAGAGCTACTAGAAAGCAATA	667		
QY	665	GCAAGAGCATCGAAACCGATAGCGAGTTGTAGAGTACGTGAGAGGATGACCCCG	724		
DB	668	ACTACGACGACAGAAACGCTGTACGGGAATTTATAGAGTACATACGCTCTCTTTGACCCAG	727		
QY	725	ACATAATAGTTCGGTTTAAACAACTATTTTCGATTCGCTTATCTTTTTCGAGGCGGCC	784		
DB	728	ACATAATAGTTCGGTTTAAACAACTATTTTCGATTCGCTTATCTTTTTCGAGGCGGCC	787		
QY	785	GCATCTTAGGCATATAAGCTTGTATGATGAGAGAGTTGGCGGAGCCCAACCACTAGCG	844		
DB	788	ACAGNATAGAGTAAAGCTCGACCTGACAGCGGTGTTGGCGAGAGCCAGTATGAGCG	847		
QY	845	TACATGGGCACTCTCTGCTCCCTGGCAGGCTTAACGTAGATCTCTATGACATATGCCGAG	904		
DB	848	TCTATGGACATGCTCTCAGTGCAGGAGTAAAGCTAAACGCTAGACCTCTACAACTACCTGGAGG	907		
QY	905	AGATGCCAGATCAAGTAAAGCTCTCGAGGAGTTCGAGAGTCTCTAGGCGGTATGA	964		
DB	908	AARTGCATGAGATTAAGGTAAAGAGCTCTCGAGGAGGTCGCCGAATACCTAGGCGGTATGC	967		
QY	965	AGAAGGTGAACGGCTTATCATCAATTGCTGGGAGATTCCAGACTATTGGGACGCCGA	1024		

968	GCAAGRCGAGCGCGTACTAATAGATGTTGGCGGATCCCGAGATTCTGGAGACGACGAGA	1027
1025	AGAAAGACCACTATTACTGCAATACGCGCGGACGATGTCCGCGTACTTTACGGCTTAG	1084
1028	AGAAACGCGCGTACTAGAGCGTTATGCCCTCGACGATGTGAGAGCCACCTACGCGCTCG	1087
1085	CCGAGAGATATTCGCGTTTGTCTATPCAGTTGTCTAGTACGTAACAGGTCTCCCACTAGACC	1144
1088	CCGAGAGATATTCGCGTTTGTCTATPCAGTTGTCTAGTACGTAACAGGTCTCCCACTAGACC	1147
1145	AGGTAGTGCATGAGTGTGGCTTTCGACTTGTATGATGTTGATGATGATGATGATGATGATG	1204
1148	AAGTCGGGCGTATCGGGCGTAGGTTTCGCTTAGAATGATGATGATGATGATGATGATGATG	1207
1205	AGATGAAGAGCTGTGTCCGAAACCGCTGTAGCGCCGAGAGAGACTTTACCGTGGCGCTA	1264
1208	ATATGAAGAGCTGTGTCCGAAACCGCTGTCAAGCGCGGAGAGAGACTTACAGGAGGAGAG	1267
1265	TAGTTCTTTAGCCGTTGAGAGCGGTGACGAGAAATATAGCCGTACTCGACTTTAGCTCGA	1324
1268	TAGTACTAAAGCCCTTAAAGGGTGTCCATGAGAACGTAGTAGTGTCTCGACTTTAGCTCAA	1327
1325	TGTACCCAAACATCATGATAAAGTACATGTTGCTCTGACACGCTTGTGAGGCTGGTG	1384
1328	TGTACCCCAACATATGATTAAGTACATGTTGGGCGCTTGACACCATTAATTGACGACCCCT	1387
1385	AAAGTGTGGCGAGT---GTGTTGCTGGGAGGCGCCGAGGCTCAAGCACAGGTTTCCGTA	1441
1388	CAGAGTGGAGAGTACAGTGGATGCTACGTAGCCCCCGAAGTCGGGCACATGTTTAGGC	1447
1442	GGTGTCCCGCGCTTCTCAAGAGCTTTTGAGAGGCTGTGTAGAGCTTGTGTAGAGCTGTG	1501
1448	GCTCGCCCTCCGCGTCTTTTAAAGACCGTGTGAGAACCTCATAGCGCTGGGTAAAGAG	1507
1502	TGCTGTGCTGTAATGAGAGATATCTCCGATAGCCCGAGATATCGATGTTGATGATAA	1561
1508	TACGTGAAGAGATGAAGAGGTTTCCCCCAGATAGCCCCAGATATACCGATATACGATGAC	1567
1562	GGCAGAGCGCTTGAAGGTTCTTGCNAACGCTAGTTACGGCTTACATGGGTTGGAGCGCG	1621
1568	GCCAGAGGCACTCAAGGTGCTAGCCAGCGTAGCTACGGCTACATGSGATGGGTGACG	1627
1622	CTAGTGTGTTTTCAGGAGTGGCAAGGCTGTCAAGCGCTTTCAGCGCTTGGGTTAGGACCTC	1681
1628	CTCGCTGTGTTCTGTAACCGCTGCGCAGAGGCTGTAAACAGCCCTGGCGCGCTAACCTG	1687
1682	GCACGCCATCAACATAGCTCTGTAAGCTCAAGTGTATCTACGCTGATCTACGCTGACACAG	1741
1688	TCTCAGCAATAGAAATATGCTAGGAAGCTCGGCGCTCAAGTAAATATACGAGACAGGACT	1747
1742	CGCTCTTCGTGACCTATGATCCGAGAGAGGTGGAAATTTTCATCAAAATTTATAAGGAGG	1801
1748	CCCTATTCTAATACCTATGATATCGAGAGGTAAAGAGCTAATAGAAATTCGTCGAGAAAC	1807
1802	AGCTGGGTTTCAAAATCAAGCTAGAGAGGTGTAAACAGCTTATCTTTTACAGAGGCTA	1861
1808	AGCTAGGCTTCGAGATAAAGATAGACAGGTATACAAAGAGAGTGTCTTTTACCGAGGCA	1867
1862	AGAGAGGTACGCTGGCTTCTCGAGGAGCGGAGCTATAGATATTCGCGTTTCGAGGCTG	1921
1868	AGAGAGGTACGCTGGCTTCTCGAGGAGCGGAGCTATAGATATTCGCGTTTCGAGGCTG	1927
1922	TACGTGGCATTTGTTGTGAACTCGCCAGAGAGGTTTCAGACTAAGTTTTCGAAATAGTAT	1981
1928	TTAGAGGCACTGTTGTGAGCTAGCTTAAGAGGTGCAAGAGAGGTAGCAGAGATATATAC	1987
1982	TGAAGCGAGTGTAGGTGAAACAGGCTGTAGAGTACGTCAAGAGAGTGTGAAAGAGTGTG	2041
1988	TGAAGCGGAGAGACATAAATAGAGCCATTAAGCTACATAAGAGAGGTGCTGTGAGAAAGCT	2047
2042	AGAGAGGCAAGTTCCTCATAGAGAGCTTGTATCTGAGAGAGCCCTTAGTBTAGGCTTGTG	2101
2048	GAGAGGCAAGATATCCCATTAACCAAGCTCGTAATATATGGAAGAGCTTTGACAAAGAGATCG	2107

FEATURES		Location/Qualifiers	
source			
1. .185300		/organism="Aeropyrum pernix"	
/mol_type="genomic DNA"		/strain="K1"	
/db_xref="taxon:56636"		103. .318	
/gene="APES066"		103. .318	
/gene="APES066"		/note="similar to PIR:D71070 percent identity:30.000 in 30aa. motif= zinc finger, C2H2 type, domain"	
/codon_start=1		/transl_table=11	
/product="71aa long hypothetical protein"		/protein_id="BAA80967.1"	
/db_xref="GI:5105655"		/translation="MRCRVVAGSKVORIGDKEDYKMLVTEGVSFVRVAIIDGMLKCPIC	
BAIFSTPRDLEVHIAHYVITLSRKNK"		complement (1004. .1744)	
/gene="APE1958"		complement (1004. .1744)	
/gene="APE1958"		/note="similar to PIR:E69068 percent identity:43.830 in 235aa."	
/codon_start=1		/transl_table=11	
/product="246aa long hypothetical NH(3)-dependent NAD(+) synthetase"		/protein_id="BAA80968.1"	
/db_xref="GI:5105656"		/translation="MLGVSGGVDSLSLALAYDAVGSGRVTLIMPDRVTPTRDVED	
ALRLVRSFGVEHAYIDISPIVMYISALPIFEDEKRDVPGVGNLRAIRANILYVYAN		KLGKLVLTGDSSEYLYGTYKDAACDVPLTVLYKQVRLAELIGVPRDIAYKP	
SSPLAKGHEAEALGLSYNEIDVILYSRFLKI PWEEIPRATGLERAKVERVRLHE		ASSHKRSPASDLGSEIKHYKQHGAKK"	
complement (1944. .3257)		/gene="APE1959"	
complement (1944. .3257)		/gene="APE1959"	
/notes="similar to OWL:AP000006150 percent identity:35.734 in 361aa."		/codon_start=1	
/transl_table=11		/product="437aa long hypothetical DNA/pantothenate metabolism flavoprotein"	
/protein_id="BAA80969.1"		/db_xref="GI:5105657"	
/translation="MEAEARGVAYRGEAVNTARRGGLNLSDVVTHPSMDILSEVNNY		IEGKTVLGVTASVALYRSLDLRLRRGARVITVMTPEAKLVSPFHFHWSGGPV	
YTGTGGEHISITARAASAVVAPATLSTLAKIAHGVDNPNVALAASVINGVKPPIA		VPAMHGMYESPQAREVVDLRSSQGLVVDPKIEGGVAKYQPDTHAVGRITAAQARKGL	
RDLEGTRALVTLGSTREWDVRFISNPSSGVNGLAEALYARGAEVDVAGYTSVE		IPIHLFTVKTETEDMAAAVEELTSKREYDAVVAAPVDVFPAGAFEGKISGGRLV	
LELEPTFKVLGTRARRPKVLVFAAEYVDNLSLRDPALEKKEKYDADLVANRVGE		GVGFAPGLDLDLMDSGEAVLKGSPHKEIVAAVIAVIADEIAKLISR"	
2755. .3078		/gene="APE1960"	
2755. .3078		/gene="APE1960"	
/note="similar to OWL:AP000006151 percent identity:39.474 in 76aa."		/codon_start=1	
/transl_table=11		/product="107aa long hypothetical protein"	
/protein_id="BAA80970.1"		/db_xref="GI:5105658"	
/translation="MLPCLAGTAITGLPYPIIDTAARATGLSTTPCAILARVESVAGA		TTWBAALAIIDMCSTFPVNYTPPEAQWNISGDTSLAASGVITVITLAPLISQRA	
RSSER"		complement (3209. .4027)	
/gene="APE1961"		complement (3209. .4027)	

FEATURES		Location/Qualifiers	
source			
1. .185300		/organism="Aeropyrum pernix"	
/mol_type="genomic DNA"		/strain="K1"	
/db_xref="taxon:56636"		103. .318	
/gene="APES066"		103. .318	
/gene="APES066"		/note="similar to PIR:D71070 percent identity:30.000 in 30aa. motif= zinc finger, C2H2 type, domain"	
/codon_start=1		/transl_table=11	
/product="71aa long hypothetical protein"		/protein_id="BAA80967.1"	
/db_xref="GI:5105655"		/translation="MRCRVVAGSKVORIGDKEDYKMLVTEGVSFVRVAIIDGMLKCPIC	
BAIFSTPRDLEVHIAHYVITLSRKNK"		complement (1004. .1744)	
/gene="APE1958"		complement (1004. .1744)	
/gene="APE1958"		/note="similar to PIR:E69068 percent identity:43.830 in 235aa."	
/codon_start=1		/transl_table=11	
/product="246aa long hypothetical NH(3)-dependent NAD(+) synthetase"		/protein_id="BAA80968.1"	
/db_xref="GI:5105656"		/translation="MLGVSGGVDSLSLALAYDAVGSGRVTLIMPDRVTPTRDVED	
ALRLVRSFGVEHAYIDISPIVMYISALPIFEDEKRDVPGVGNLRAIRANILYVYAN		KLGKLVLTGDSSEYLYGTYKDAACDVPLTVLYKQVRLAELIGVPRDIAYKP	
SSPLAKGHEAEALGLSYNEIDVILYSRFLKI PWEEIPRATGLERAKVERVRLHE		ASSHKRSPASDLGSEIKHYKQHGAKK"	
complement (1944. .3257)		/gene="APE1959"	
complement (1944. .3257)		/gene="APE1959"	
/notes="similar to OWL:AP000006150 percent identity:35.734 in 361aa."		/codon_start=1	
/transl_table=11		/product="437aa long hypothetical DNA/pantothenate metabolism flavoprotein"	
/protein_id="BAA80969.1"		/db_xref="GI:5105657"	
/translation="MEAEARGVAYRGEAVNTARRGGLNLSDVVTHPSMDILSEVNNY		IEGKTVLGVTASVALYRSLDLRLRRGARVITVMTPEAKLVSPFHFHWSGGPV	
YTGTGGEHISITARAASAVVAPATLSTLAKIAHGVDNPNVALAASVINGVKPPIA		VPAMHGMYESPQAREVVDLRSSQGLVVDPKIEGGVAKYQPDTHAVGRITAAQARKGL	
RDLEGTRALVTLGSTREWDVRFISNPSSGVNGLAEALYARGAEVDVAGYTSVE		IPIHLFTVKTETEDMAAAVEELTSKREYDAVVAAPVDVFPAGAFEGKISGGRLV	
LELEPTFKVLGTRARRPKVLVFAAEYVDNLSLRDPALEKKEKYDADLVANRVGE		GVGFAPGLDLDLMDSGEAVLKGSPHKEIVAAVIAVIAVIADEIAKLISR"	
2755. .3078		/gene="APE1960"	
2755. .3078		/gene="APE1960"	
/note="similar to OWL:AP000006151 percent identity:39.474 in 76aa."		/codon_start=1	
/transl_table=11		/product="107aa long hypothetical protein"	
/protein_id="BAA80970.1"		/db_xref="GI:5105658"	
/translation="MLPCLAGTAITGLPYPIIDTAARATGLSTTPCAILARVESVAGA		TTWBAALAIIDMCSTFPVNYTPPEAQWNISGDTSLAASGVITVITLAPLISQRA	
RSSER"		complement (3209. .4027)	
/gene="APE1961"		complement (3209. .4027)	

```
/db_xref="GI:5105663"
/translation="MLAVVAGITSVEVRLMVDVIVIKLRLKRLGSOYSEQVD
KWIPTFATIREGLKSIAEKLEBRLISASNEACVEVLVNMIMGMDENDAIGSLRG
SLAAVRGVVSRLCNKKALDSSI"
/complement(8023..8352)
/gene="APE1966"
/complement(8023..8352)
/codon_start=1
/transl_table=11
/product="272aa long hypothetical stress protein"
/protein_id="BAA80971.1"
/db_xref="GI:5105659"
/translation="MPLFTIDYNDPFIQIKTERISAVGLGTWAIIRDYSSALESVYVA
IERGINHIDTAENYAGAEAEFVGRVGRVDRVYVTKLPPFRPDDEKVAARA
SARMGVSVYDITLHWPDOPOTPIKQVKSLSAIAEAGLARVYGVNFKSBOLEALIT
ATKHEIVNQYKSVLDRAELIPLAVEKSLILEAYSPERGHVAEVLAKIGRQ
YKTPQVVALFVSRNFIALPFAERREHYDELGSIGWRLKPEDIELLEKLI"
/100..5428
/gene="APE1962"
/100..5428
/gene="APE1962"
/100..5428
/notes="similar to OWL:AB00952240 percent identity:53.027
in 413aa."
/codon_start=1
/transl_table=11
/product="42aa long hypothetical serine
hydroxymethyltransferase"
/protein_id="BAA80972.1"
/db_xref="GI:5105660"
/translation="MDGLKELDSLDPMSVLKIAELTTRHNVRRLRESINLIASENMV
SLTAKAYLSDFMFRYAEAGKPFKRYQGTTRYIDLEVLTELMGSMGNTLVELRPVS
GTIANASVFLVAPGDKAVIAPVQAGAHVSHTKFGLGALGIEQVSMFYDEENMVD
VDKAVLIEBVKPRFAVLGGSVVIFPHPTREIAEAIHVSQGLIYDAAHVLIIMGGA
WPNPLEKADAVGSHKTFPGQGGAVFRDQLYKYSKTIFFPWNVSNHHLIRIPA
TALTAVMKUIGRDYASQVTSNARKLABALAAAGLVYGEHLGYTHSHQVVDVRLG
GGAKCSLSEESINVNKLLPMDPPPEAVPGSIRIGVQEVNLTGKMGEMEEIAKL
IRKVLIDKDFKVAEQVKEFRKQFMKIHYSLDDKQVRIDITDIADLSLAQ"
/5736..6935
/gene="APE1963"
/5736..6935
/notes="similar to PIR:H69252 percent identity:34.711 in
363aa. motif=iron-containing alcohol dehydrogenases
signatures"
/codon_start=1
/transl_table=11
/product="399aa long hypothetical alcohol dehydrogenase"
/protein_id="BAA80973.1"
/db_xref="GI:5105661"
/translation="WMFSLGPRLVDFVPRRIYVNESLPRALETLIGDKGVKHLVW
TDGKVAAMSPKBAEHSVAGISGVFCFDGVTPEPDFVDGDAIASEARRCRBAIVAV
GGSVIADAKAGFKFVRPDADLGLAPENVLGDSGIVLVAFTPTSGTSYASGI
VLTKRVNGREKIAVGSYEVVPIYASILDPSITLGNPERLIVGTAVDALAHSEVIAST
NANPLSALAAKAAEIVFRLPEVVDGNYDARAEMHIAATWAGMAFTNSGLGASH
IAHPLGARLGHGTGATVGLPHVVRNFSRSDYARKYWEKLLKLLSLGLBERGSL
ADHLEELYARVQTRVELAGIERERYLTIADEVAAAFRDPPELAFNPNVPTMEQYK
TLLYNMV"
/5961..7371
/gene="APE1964"
/5961..7371
/gene="APE1964"
/codon_start=1
/transl_table=11
/product="136aa long hypothetical protein"
/protein_id="BAA80974.1"
/db_xref="GI:5105662"
/translation="MDAMVQPRFPTSKEWAEFFCKTLDSSEKYSRAAKGWVHPILFKIS
DGGQGFILDLKDGRCGLGVQWPDFDTSRADPVIIISRLSDMLDVINGKVNLTALLTRR
LRIKGDLSLILRYSABAAMVSAQRVGVSGQQ"
/complement(7482..7856)
/gene="APE1965"
/complement(7482..7856)
/gene="APE1965"
/codon_start=1
/transl_table=11
/product="124aa long hypothetical protein"
/protein_id="BAA80975.1"
```

```
/db_xref="GI:5105663"
/translation="MLAVVAGITSVEVRLMVDVIVIKLRLKRLGSOYSEQVD
KWIPTFATIREGLKSIAEKLEBRLISASNEACVEVLVNMIMGMDENDAIGSLRG
SLAAVRGVVSRLCNKKALDSSI"
/complement(8023..8352)
/gene="APE1966"
/complement(8023..8352)
/codon_start=1
/transl_table=11
/product="272aa long hypothetical stress protein"
/protein_id="BAA80971.1"
/db_xref="GI:5105659"
/translation="MPLFTIDYNDPFIQIKTERISAVGLGTWAIIRDYSSALESVYVA
IERGINHIDTAENYAGAEAEFVGRVGRVDRVYVTKLPPFRPDDEKVAARA
SARMGVSVYDITLHWPDOPOTPIKQVKSLSAIAEAGLARVYGVNFKSBOLEALIT
ATKHEIVNQYKSVLDRAELIPLAVEKSLILEAYSPERGHVAEVLAKIGRQ
YKTPQVVALFVSRNFIALPFAERREHYDELGSIGWRLKPEDIELLEKLI"
/100..5428
/gene="APE1962"
/100..5428
/gene="APE1962"
/100..5428
/notes="similar to OWL:AB00952240 percent identity:53.027
in 413aa."
/codon_start=1
/transl_table=11
/product="42aa long hypothetical serine
hydroxymethyltransferase"
/protein_id="BAA80972.1"
/db_xref="GI:5105660"
/translation="MDGLKELDSLDPMSVLKIAELTTRHNVRRLRESINLIASENMV
SLTAKAYLSDFMFRYAEAGKPFKRYQGTTRYIDLEVLTELMGSMGNTLVELRPVS
GTIANASVFLVAPGDKAVIAPVQAGAHVSHTKFGLGALGIEQVSMFYDEENMVD
VDKAVLIEBVKPRFAVLGGSVVIFPHPTREIAEAIHVSQGLIYDAAHVLIIMGGA
WPNPLEKADAVGSHKTFPGQGGAVFRDQLYKYSKTIFFPWNVSNHHLIRIPA
TALTAVMKUIGRDYASQVTSNARKLABALAAAGLVYGEHLGYTHSHQVVDVRLG
GGAKCSLSEESINVNKLLPMDPPPEAVPGSIRIGVQEVNLTGKMGEMEEIAKL
IRKVLIDKDFKVAEQVKEFRKQFMKIHYSLDDKQVRIDITDIADLSLAQ"
/5736..6935
/gene="APE1963"
/5736..6935
/notes="similar to PIR:H69252 percent identity:34.711 in
363aa. motif=iron-containing alcohol dehydrogenases
signatures"
/codon_start=1
/transl_table=11
/product="399aa long hypothetical alcohol dehydrogenase"
/protein_id="BAA80973.1"
/db_xref="GI:5105661"
/translation="WMFSLGPRLVDFVPRRIYVNESLPRALETLIGDKGVKHLVW
TDGKVAAMSPKBAEHSVAGISGVFCFDGVTPEPDFVDGDAIASEARRCRBAIVAV
GGSVIADAKAGFKFVRPDADLGLAPENVLGDSGIVLVAFTPTSGTSYASGI
VLTKRVNGREKIAVGSYEVVPIYASILDPSITLGNPERLIVGTAVDALAHSEVIAST
NANPLSALAAKAAEIVFRLPEVVDGNYDARAEMHIAATWAGMAFTNSGLGASH
IAHPLGARLGHGTGATVGLPHVVRNFSRSDYARKYWEKLLKLLSLGLBERGSL
ADHLEELYARVQTRVELAGIERERYLTIADEVAAAFRDPPELAFNPNVPTMEQYK
TLLYNMV"
/5961..7371
/gene="APE1964"
/5961..7371
/gene="APE1964"
/codon_start=1
/transl_table=11
/product="136aa long hypothetical protein"
/protein_id="BAA80974.1"
/db_xref="GI:5105662"
/translation="MDAMVQPRFPTSKEWAEFFCKTLDSSEKYSRAAKGWVHPILFKIS
DGGQGFILDLKDGRCGLGVQWPDFDTSRADPVIIISRLSDMLDVINGKVNLTALLTRR
LRIKGDLSLILRYSABAAMVSAQRVGVSGQQ"
/complement(7482..7856)
/gene="APE1965"
/complement(7482..7856)
/gene="APE1965"
/codon_start=1
/transl_table=11
/product="124aa long hypothetical protein"
/protein_id="BAA80975.1"
```



```
QY 926 AGAGTCTCGAGAGTTCGACGATATCTAGCGGTGATGAAGAAGTGAACGCGTTATCA 985
Db 89528 AGACTCTCGAGAGTTCGCGACTACCTAGCGGTGATGAAGAAGTGAACGCGTTATCA 89469
QY 986 TCAATTGTGGGAGATTCACGACTATTCGAGACGCCCGAAGAAGACCACTATTACTGC 1045
Db 89468 TAGAGTGTGGCAGATAGGGAGTACTGGGATGACCCCTCAAGCGGAGATATCTCAGGA 89409
QY 1046 AATACGCGCGACGATGTCGCGCTACTTACGGCTTACGCGAGAGATATTCGCGTTTG 1105
Db 89408 AGTACCTAAGGAGACGATGTCAGAGCACCATGCGGCTGCGCAGAGAATTCCTACCCCTCG 89349
QY 1106 CTATCCAGTGTGCTAGTAACAGGTCTCCCACTAGACAGGTAGGTGCGATAGTGTG 1165
Db 89348 GGGCGGAGCTCTCAAGTCTCAGGGCTCCCGCTGACAGGATATGCGCGCCACGCGTAG 89289
QY 1166 GCTTTCGACTTGAATGTTACTGATACGCGCGCGCTTTAAGATGAAGAGCTTGGCCGA 1225
Db 89288 GCTTCAGACTGGAGTGGAGGCTCATACGCGAGCGGCTAAGCTCGCGAGCTTGGTCCCA 89229
QY 1226 ACCGGTGTGAGCGCCGACAGAGACTTACCGTGGCGCTATAGTCTTGAGCGGTGAGAG 1285
Db 89228 ACAGGTTGAGAGGACGAGGCGCAGTACGCTGGCGCCATAGTACTCAGGCCCAAGCCGG 89169
QY 1286 GCGTGCACAGAAATATAGCCGCTACGCTTGTAGTTCGATGACCAACATCATGATPAA 1345
Db 89168 GGGTTACAGAGGATATGCGCTCTGGAATTCGCAAGCATGTACCCGGAATATCATGTTAA 89109
QY 1346 AGTACAACTTGGTCTGACACGCTTGTGAGCGCTTGTGAAGGTGCGCAGTGTGTT 1405
Db 89108 AGTATAAGTGGCGCCGACATCTGCTCAGCGCTGGGAGGAGTATGGAGGAGGAGG 89049
QY 1406 GCTGGAGCGCCCGAGGTCACAGCAAGGTTCGTTAGGTGTCGCGCGCTTCTTCAAGA 1465
Db 89048 TATACAGGCTCCGAGGTGGGCCACAGTTCAGGAGAGCCCTCCCGCTTCTTCAAGA 88989
QY 1466 CAGTCTTGAGAGGCTGTAGAGCTTCGPAAGCTGTGCGTGTGCTGAAATGAAGATATC 1525
Db 88988 AAATACTTGAGAGGTTCCTAAGCTGGAGGAGGAGATTAAGGAGGAGATGAAGAGCATC 88929
QY 1526 CTCGGATAGCCAGAAATATGACTTGTGATGAAGGAGGAGGAGGCTTGAAGGTTCTTG 1585
Db 88928 CGCCGACAGCCGCGAGTACAGCTCTCGACGAGGAGGAGGAGGAGGAGGAGGAGG 88869
QY 1586 CAAAGCTAGTACGCTACATGCTGTGAGCGGCGCTAGGTGTTATTCAGGAGTGGC 1645
Db 88868 CAAAGCTAGTACGCTACATGCTGTGAGCGGCGCTAGGTGTTATTCAGGAGTGGC 88809
QY 1646 CAAAGCTAGTACGCTACATGCTGTGAGCGGCGCTAGGTGTTATTCAGGAGTGGC 1705
Db 88808 CAAAGCTAGTACGCTACATGCTGTGAGCGGCGCTAGGTGTTATTCAGGAGTGGC 88749
QY 1706 AACTAGGCTCAAGGTGATCAGGTCGACAGATTCGCTCTTCGTCGACCTATGATCCGG 1765
Db 88748 AGCTAGGCTCAAGGTGATCAGGTCGACAGATTCGCTCTTCGTCGACCTATGATCCGG 88689
QY 1766 AGAGGTGGAATTCATCAAAATTAAGAGAGAGGCTGGGTTGGAATCAAGCTAG 1825
Db 88688 AGAGGTGGAATTCATCAAAATTAAGAGAGAGGCTGGGTTGGAATCAAGCTAG 88629
QY 1826 AGAGGTGGAATTCATCAAAATTAAGAGAGAGGCTGGGTTGGAATCAAGCTAG 1885
Db 88628 AGAGGTGGAATTCATCAAAATTAAGAGAGAGGCTGGGTTGGAATCAAGCTAG 88569
QY 1886 AGAGGTGGAATTCATCAAAATTAAGAGAGAGGCTGGGTTGGAATCAAGCTAG 1945
Db 88568 AGAGGTGGAATTCATCAAAATTAAGAGAGAGGCTGGGTTGGAATCAAGCTAG 88509
QY 1946 CCAAGAGGTTCAAGTATAGGTTGTCGAAATAGTATGAAGAGGAGTGGAGTGAACAGG 2005
Db 88508 CCAAGAGGTTCAAGTATAGGTTGTCGAAATAGTATGAAGAGGAGTGGAGTGAACAGG 88449
```

```
QY 2006 CTCTAGAGTACGTCCAGAGATTTGTAAGAGTTGGAGAGGCGCAGGTTCCTCATAGAGA 2065
Db 88448 CTCTAGAGTACGTCCAGAGATTTGTAAGAGTTGGAGAGGCGCAGGTTCCTCATAGAGA 88389
QY 2066 AGCTTGTAACTCGAAGACCCCTTAGTAAGCGTCTTGAGAGGTACACACGAGGACCCAC 2125
Db 88388 AACTGTTCATTTGGAAGACTCTGCACAGCGCTCCATCAATGTACGAGGCGGAGCAGCCC 88329
QY 2126 ACCTGCTTGAGCGAAGAGATGCTGTACAGAGGCTACCGGGTAAACCCAGGCGCAAGA 2185
Db 88328 ACCTGACCGCAGCGCTACTTATGAGAGGCGTGCATAAAGGTCGAGGCGCAGGCGTAA 88269
QY 2186 TAGGTATGTAATAGTCAAGGCTGTCGCGGTATCAGTCAAGAGCATGSCCTACTTCA 2245
Db 88268 TCAGCTACGCTGTCGACCAAGGCTCAGGCGCTCTATACGAGGCGGAGCCTTATTTTA 88209
QY 2246 TGCTCAAGGATCTCAGCAGATAGAGTACCTTCTGACTT 2390
Db 88208 TGCTAGCA--GGAGGAGTTGACGTGGAGTACTATGTCGACAGCAGGTGTGTCGAG 88152
QY 2306 CTGATTTGAGAACTCTGGGCTACTTTGGCATCACCGAGAAAGCTGAAAGCAAGTGCAA 2365
Db 88151 CGCTCTGCGCATCTCCAGTACTTCGAGTAACTGGAAGAGCTGTAAGGAGGAGGAGAA 88092
QY 2366 CTGGCAGAGAGCTCTCTTCGACTT 2390
Db 88091 ---GGCAGAGCAGCTCTCGACTT 88070
```

```
RESULT 11
AB017501
LOCUS AB017501 2319 bp DNA linear BCT 06-OCT-1999
DEFINITION Aeropyrum permix gene for DNA polymerase II, complete cds.
ACCESSION AB017501
VERSION AB017501.1 GI:4519583
KEYWORDS DNA polymerase II.
SOURCE Aeropyrum permix
ORGANISM Aeropyrum permix
Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
Desulfurococaceae; Aeropyrum.
REFERENCE
1 (sites)
Cann, I.K., Ishino, S., Nomura, N., Sako, Y. and Ishino, Y.
Two family B DNA polymerases from Aeropyrum permix, an aerobic
hyperthermophilic crenarchaeote
J. Bacteriol. 181 (19), 5984-5992 (1999)
JOURNAL
MEDLINE 59429845
PUBMED 10498710
REFERENCE
2 (bases 1 to 2319)
Ishino, Y. and Cann, I.K.
Direct Submission
Submitted (04-SEP-1998) Yoshizumi Ishino, Biomolecular Engineering
Research Institute, Department of Molecular Biology; 6-2-3
Furuedai, Suita, Osaka 565-0874, Japan (E-mail:ishino@beri.co.jp,
Tel:+81-6-872-8203(ex.453), Fax:+81-6-872-8219)
FEATURES
Location/Qualifiers
1..2319
/organism="Aeropyrum permix"
/mol_type="genomic DNA"
/strain="K1"
/db_xref="taxon:56636"
1..2319
/gene="polB"
1..2319
/gene="polB"
/codon_start=1
/product="DNA polymerase II"
/protein_id="BA075663.1"
/db_xref="GI:4519584"
/translation="MRGSTVIIIMRGADGSRVVVYGFYFVLPDGSVGLDQL
AAMIRLSRSPFISVRRRFGVEALKVTLLVPASVREYRVRRLGGVRDV
LKADIPALRFIIFDNLPMRVAVRVPVPHGYSVDRATLSGDIRDETRIOED
PLKGLRVMAFDIEVYSKQRTDPKDPVIMIGLQAGGEITILEABDRSDKVIAGFV
```


BRVKSIDPDIIVGONRFDWPLVERVLGVKLAIVGRSVEPQELGYHSVSGRL
NVLLDFABELHEVUKVTLSEVADYLGVEYWDPSKELRLKXLLDDVRSWGLAE
KFLPFGQLSQVSGSLPDLQVMAASVGLRLEWRLIREAKLGLVLPKRVSEBRYAGA
IVLRPKFQSHEDJAVLDFASMIPIVNVKINWPDLLVRPGEYGESEVYTAPEVGHF
RKPPGPFKILERFLSWMRSIRSEMKHPDPSFYKLLDEROKAKLLANASVYMG
WPHARWYCRCAEAVTANGRSIIIRTAIRKAGELGLEVIYDGLDLSLFDKNDPEKVERLI
RFVEBELGFDIKVDKVVRRVFFTEAKRYVGLVDGKIDVVGFEAVRGDWSBELAKETQ
PKVAEIVLTKGSDVEADVVRNLIIEKLRCGVDMRKLVIWKTITRPPSVARQPHVT
ALLMERAGIKVBEPAKIGVYVTKGSPILYTRAKPYFPMASKEEVDVEYVDKQVPA
SAHTSSNFVIEKPLKGGGRQSLDDPMRKG*
BASE COUNT 557 a 531 c 802 g 429 t
ORIGIN

Query Match 30.58; Score 735.2; DB 1; Length 2319;
Best Local Similarity 59.14; Pred. No. 7.3e-103; Indels 48; Gaps 5;
Matches 1375; Conservative 0; Mismatches 903;
Qy 65 TAATCATATGGGGTATTCCTGAGAACGCGAGAGGGTAGTCTCATTTGACAGGTCTTTTC 124
Db 23 TCATACTCTGGGGCGGGGCTGACGGTCTAGGGTTGCTGTTCTACGGCGAGTTCA 82
Qy 125 GCCCATACTTCTATGCGTGTCTTCACGGGCGCGGATCTTAAGCAGGTAGCAACGTA 184
Db 83 GACCTTACTTCTAGCTGTCTGCGGACGCGGAGCGTAGGCTCGACACAGCTGGCCTATGA 142
Qy 185 TTCTGTGATTCAGTAGGACCAAGAGCCGATTATAGTGTAGAGGATGACAGAGGAAGT 244
Db 143 TAAGGAGCTCTCAAGGCGGTCCAGCCCAATACTCTCGTTGAGAGGGTTGACGCGAGGT 202
Qy 245 ACTTCGGAGGCTCTGAGGCTTCTACGTATTCGCACCGCTGCTACCGAGGCTGTTAGGG 304
Db 203 TTATAGGCGAGGAGTGGAGGCTCTCAAGGTAAACCCCTCGTCCCGCAAGGCTCAGGG 262
Qy 305 AGTATCGGAACCTGTAAGAACGTTGATGGTGTAGGATGTCTAGAGCGCGATATAC 364
Db 263 AGTATAGGAGGCTGTGCTAGGCTGGGGGCGGTAGGGATGTCTGAAGGCGCGATATAC 322
Qy 365 GCTTCGCTATCGCTATCTCATAGATCACGATCTATTTCTTCCACCTGTGTACCGTGTAG 424
Db 323 CCTTCGCGCTCAGGTTCTATAGATCTTAACCTATACCTATAGATGTGTGTGCGCG 382
Qy 425 AGGCTGAGCCCTCGAGAACAGATGGGCTTCCGTGTGACAGGATATACCTGTTAAGA 484
Db 383 AGGTTAGGAGGCTCGGCTCCCGACGCTATACGTTGACAGGCTACACGCTCTCCG 442
Qy 485 GCAGGCGGACCTTATATGTTAGGCTCTGCGACCAACCAAGCTTCCGATCTTAGGA 544
Db 443 GC-GATATAAGGAGGACGAGACGAGGATCCAGGAGGACCC--CCTGAAGGGGCTCAGGG 499
Qy 545 TACTCGGCTTCGATTTGAAGTTTATAGCAAGCAAGGCTCGCGCTCCAGAGCGGATC 604
Db 500 TTATGGCTTTCGACATAGAGTCTACAGCAGATAGAGGACGCCAGACCCCAAGAGACC 559
Qy 605 CTGTAATAGTATAGCTGTGAAGACTGACGATGCGGATAGGTTGCTATTCATTGCAAGG 664
Db 560 CGGTAATCATGATAGGTTGACGAGCGCGAGGGAGATCGAGATCCTGAGGCGGAGG 619
Qy 665 GCAAGAGGATCGAAACCCGATACCGAGTTGTAGAGTACGTGAAGAGGTATGACCCCG 724
Db 620 ATAGGAGTGAAGAAGGTTATACCGGCTTCGTCGAGCGGCTGAAGAGTATCGACCCCTG 679
Qy 725 ACATAATAGTCCGTTTATAACAACATCATTTTCGATTCGCTTATCTTTTGGGCGCGCCC 784
Db 680 ACGTATAGTCCGCTATACCGAAGCAGGTTTCGACTGCGCATACCTCGGTTGAGAGGCCA 739
Qy 785 GCATCTAGGATTAAGCTTGTATGACTAGAGAGTTGGCGCGAGCCACCACTAGCG 844
Db 740 GAGTTCCTCGGCTGAGAGCTTCGGTCCGGAGGAG--GAGGCTGAACCCCGCCGCT 796
Qy 845 TACATGGGCAAGCTCTCTGCTCCCTCGCGGCTTAAAGTAGATCTGTAAGCTATGCGGAG 904
Db 797 TGTACGGGCACTACTCTCGCTCTCTGGCAGGCTCAACGCTAGACCTCTCTAGACTTTTGGGAGG 856

Qy 905 AGATGCCAGAGATCAAGATAAAGAGTCTCGAGAGGTCGCGAGGATCTAGGCGGTGATCA 964
Db 857 AGCTCCACAGGTTTAAAGTTTAAAGATCTCGAGGAAGTCCGCGACTACTTAGGGGTGCTGA 916
Qy 965 AGAAGAGTGAACCGCGTTATCATCAATTTGGTGGGAGATTTCCAGACTATTGGGACGACCCGA 1024
Db 917 GG-----GAGTACTGGGATGACCGGT 937
Qy 1025 AGAAGAGACCATTTATCTAGTCAATACGCGCGGAGAGTGTCCGCGCTACTTACGCGCTTAG 1084
Db 938 CAAAGCGGAGATACCTCAGGAAGTACTCAAGGAGAGATGCTCAAGAGCACCATATGGGCTGG 997
Qy 1085 CCGAGAAGATATTGCCGTTTGTATCCAGTTCCTGCTAGCTACAGTACAGCTCTCCCACTAGACC 1144
Db 998 CAGAGAGGTTCTACCTTCGGGGCGAGCTCTCAAGTCTCAGGGTCTCCCGCTGGAGCC 1057
Qy 1145 AGGTAGTCCGATAGTGTGGCTTTTCGACTTTGAATTTGATCTGATACGCGCGGCTTTA 1204
Db 1058 AGGTTATGCGCGCGAGCGTAGGCTTCAGACTGGAGTGGAGGCTCATCATGGGAGCGGCTA 1117
Qy 1205 AGATGAAGAGCTTGTCCGACCGGCTTGGAGCGCCGAGAGAGACTTACCGTGGCGCTA 1264
Db 1118 AGCTCGCGAGCTGTGCGCCAAACAGGGTTGAGAGAGCGGAGGAGTACGCTGGGGGCA 1177
Qy 1265 TAGTCTTCGAGCGGTTGAGAGCGGTGACAGAGATATAGCCGTACTCTGACTTTAGCTCGA 1324
Db 1178 TAGTACTCAGGCGCCAAAGCGGGGTTCAAGAGATATTGCCCTCTGACTTTCGAGCA 1237
Qy 1325 TGTACCCAAACATCATGATATAAGTACAAATTTGTTGCTCTGACACGCTTGTGAGGCTGTG 1384
Db 1238 TGTACCCGAATATCATGTTAAAGTATAACGTCGCGCCCGACACTCTGTTGTCAGGCTGGGG 1297
Qy 1385 ABAAGTGTGCGGAGTGTGTTGCTGGAGGCGCGGAGGTCAAGCACAGGTTCCGTAGGT 1444
Db 1298 AGAGTGTGAGAGGAGGAGGTATACAGCGCTCCGAGGTTGGGCGACACAGTTTCAGAGA 1357
Qy 1445 GTCCGCGCGGCTTCTTCAAGACAGTCTTCTGAGAGGCTGTAGAGCTTCTGAAGCGGTGC 1504
Db 1358 GCGCTCCCGCTTCTTCAAGAAATACTTTGAGAGGTTCTTAAGCTGGAGAGGCGAGATTA 1417
Qy 1505 GTGCTGAATAGAAAGTATCTCCGATAGCCCGAGATATCGACTGTTTGGATGAAGGC 1564
Db 1418 GGAGCGAGATGAAGAGCATCCGCGCAGCAGCCCGAGTACAAGCTCTCTCGAGAGAGC 1477
Qy 1565 AGAAGCGTGTAGGTTCTTTCGAAACGCTAGTATACGCTACATGGTTGGAGCGGCGCTA 1624
Db 1478 AGAAGCGGATAAGCTCTTGGCAACCGCAGCTACGCTACATGGGCTGCGCCCAATGCCA 1537
Qy 1625 GGTGGTATTTCAGGAGTGGCAAGGCTGTCAAGGCTTGGGCTAGGCACTCATACGCA 1684
Db 1538 GGTGGTACTGTAGAGAGTGTGAGAGGCGGTTTACAGCGTGGGCTAGGAGCATTAAGA 1597
Qy 1685 CCGCATCAACATAGTCTGTAACTTAGGCTCAAGGTGATCTACGCTGACAGATTCGC 1744
Db 1598 CTGCTATAAGGAGGCTGCGAGCTAGGCTGGAGGTTATATATATGAGACACGCAACGCG 1657
Qy 1745 TCTTCGTGACCTATGATCCGAGAGGTCGAAATTTTCATCAAAATTAATAAGGAGGAGC 1804
Db 1658 TCTTCGTCAAGAACGATCTGAGAAAGGTGGAAGGCTGATAAGGTTTTCGAGGAGGAGC 1717
Qy 1805 TGGGTTCCGAATCAAGCTTAGAAGGTTGTAACAACGCTTATCTTTTACAGAGGCTAAGA 1864
Db 1718 TTGGGTTCAATTAAGGTGGACAGGTTGTAACAAGGCTGATCTTTCACCGAGGCGCAAGA 1777
Qy 1865 AGAGGTAACGCTGCGCTTCTCGAGGACGAGCTATAGATATGTCGCTTTCGAGGCTGTAC 1924
Db 1778 AGAGGTATGTCGGGCTGACGCTGGACGCAAGATAGATGCTAGTTCGCTTTCGAGGCTGTCA 1837
Qy 1925 GTGGCGATTTGCTGTGACTCGCCAGAGGAGTTTCAGCTACAGGTTGTCGAAATAGTATGA 1984
Db 1838 GGGGCGACTGAGCGAGCTCGCAGAGGAGACGACGCTTCAAGGTGGCAGAGATAGTCTTAA 1897
Qy 1985 AGACGAGTGTAGTGAACAGGCTGTAGAGTACGTCAGGAAGATTTGTGAAGAGTTGGAGG 2044

```

1898 AGACCGTAGCTGGACGAGGCTGTCGACTACCTCAGGAATATAATAGGAAGCTTAGGA 1957
2045 AGGGAGGTTCCCATAGAGAGCTGTGTAATCTGAGAGCCCTTAGTAAGCTCTTGAGG 2104
1958 GGGGCGAGGTGACATGAGAACTGGTCAATTTGGAAGACTCTGACAGGGCTCCATCA 2017
2105 AGTACACAGGAGGACACACAGCTGTTGACAGCAAGAGAGTGTCTGACAGAGCTACC 2164
2018 TGTACAGGAGGAGGAGCCACGCTGACCGAGCGCTACTTATGGAGGGGCTGGCATAA 2077
2165 GGTAAGCCAGCGACAGATAGAGTATGTAATAGTGAAGGTTGGTGGCCGTATCAGTC 2224
2078 AGTGCAGCCAGGGCTAAGATCGGTGCTGACCAAGGGGTGACGGCCCTCTATA 2137
2225 AAGACATGGCCATACCTTCAATGCTCAAGATCTTAGCCAGATAGACGTGACTACTTG 2284
2138 CGAGGGCGAAGCTTATTTATGGTAGCAAGAGAGAGGTGACGTGGAGTACTATGTCG 2197
2285 TTGACACCAAAATCATCCCGCTGCTATGAGATACTGGGCTACTTTGGCATCACCAGA 2344
2198 ACAGAGAGTGGTGCAGCGGCTTCTGGCATACTTCCAGTAACCTTCGGAGTAATTGGA 2257
2345 AGAAGCTGAAGCAAGTGCACCTGGGCAAGAGTCTCTCTCGACTT 2390
2258 AGAGGCTGAAGGAGGAGGAA---GGCAGAGCACGCTCCTCGACTT 2300

RESULT 12
E36726
LOCUS Novel DNA polymerase. 2319 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel DNA polymerase.
ACCESSION E36726
VERSION E36726.1 GI:18626542
KEYWORDS JP 2000197488-A/4.
SOURCE unidentified
ORGANISM unidentified
REFERENCE
1 (bases 1 to 2319)
Kahn, A., Ishino, Y. and Sako, Y.
Novel DNA polymerase
Patent: JP 2000197488-A 4 18-JUL-2000;
BIOMOLECULAR ENG RES INST
OS aeropyrum pernix K1
PN JP 2000197488-A/4
PD 18-JUL-2000
PF 06-JAN-1999 JP 1999000956
PR AISAAC KAHN, YOSHIZUMI ISHINO, YOSHIIHIKO SAKO
PC C12N15/09, C12N9/12, C12Q1/68// (C12N15/09, C12R1:01), (C12N9/12,
PC C12R1:01),
PC C12N15/00, (C12N15/00, C12R1:01)
CC
FH Key Location/Qualifiers
FT source 1..2319
FT /organism="aeropyrum pernix K1".
FEATURES
source
1..2319
Location/Qualifiers
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 557 a 531 c 802 g 429 t
ORIGIN
Query Match 30.5%; Score 735.2; DB 6; Length 2319;
Best Local Similarity 59.1%; Pred. No. 7.3e-193;
Matches 1375; Conservative 0; Mismatches 903; Indels 48; Gaps 5;
QY 65 TAATCATATGGGTTATGCTGAGAACGGCAGAGGCTAGTCTCATTCACAGGCTTTTC 124
Db 23 TCATACTCTGGGGGGGGGGGCTGACGGCTAGGGTTGTCTGTCTACGGCGAGTTCA 82
QY 125 GCCCATCTTCTATGCGCTGCTTGCAACCGGGCGCGATCCTTAAGCAGGTAGCACAACTGA 184

```

```

83 GACCTTACTTCTAGCTGCTCCCGACGGAGCGTAGGCCTCGACCAGCTGCGAGCTATGA 142
185 TTCTGTCATTAGTAGGCCAAAGAGCCGATTTAGGTGTAGAGGATGACAAGAGGAAGT 244
143 TAAGAGGCTCTCAGAGCCGCTCCAGCCCAATACTCTCGGTTGAGAGGGTTGACGCGAGT 202
245 ACTTCGGAGAGCCTCTAGGGTCTTACGTATTGCGACCGTCTACCGAGGCTGTTAGGG 304
203 TTATAGCAGGGAGGTGGAGGCTCTCAAGGTAACACCCTCGTCCCGCAAGCGTCAAGG 262
305 AGTATCGCGAAGCTCGTAAAGAACGTTGATGGTGTGAGGATGTTCTAGAGCGCGATATAC 364
263 AGTATAGGAGGCTGTGCTAGGCTGGGGGCGTTAGGATGTCTCTCAAGGCGGATATAC 322
365 GTTTCGCTATGCGCTATCTCATAGATCACGATCTATTTCCTTTTCCCTGATACCGTGTAG 424
323 CTTTCGCCCTCAGGTTTCATCATAGATTTAACTTATACCTTATACCTTATGAGATGATGTG 382
425 AGGCTGAGCCCTCGAGAAAGAATGGCTTCGCTGTCGACAGAGTATACCTGTTAAGA 484
383 AGGTTAGGAGGTCGCGCTCCCGCATATAGCGTTGACAGGGCTTACACGCTCTCCG 442
485 GCAGSCCGAGCCACTTTATGCTGAGGCTCTCGCACCAACCAAGCTTCCCGATCTTAGGA 544
443 GC-GATATAAGGAGGACGACAGAGATCCAGAGGACCC--CCTGAGGGCTCTCAGG 499
545 TACTCGCTTCGATATTGAAGTTTATAGCAAGAGGTCGCGGTCGCGGTCCAGAGCGCGATC 604
500 TTATGGCTTTCGACATAGAGGCTTACAGCAAGATGAGGACGCCAGACCCCAAGAGGACC 559
605 CTGTAATAGTATGACTGTGAGACTGACGATGGCGATGAGTGTCTATTTCAGAGG 664
560 CGGTAATCATGATAGGGTTGACAGAGCGCGAGGAGGATCGAGATCTCTGAGGCGGAGG 619
665 GCAAGAGCATCGAAAACCGATACCGGAGTTTGTAGAGTACGTGAAGAGGTATGACCCCG 724
620 ATAGGAGTGACAAAGGTTTATAGCGGCTTCTGTCGAGCGGTGAAGATGACGACCTG 679
725 ACATAATAGTCGGTTATACAAATCAATTCGATGAGTGGCTTATCTTTGAGGCGCGCC 784
680 ACGTATAGTCGGCTATAACCAAGACAGGTTTCGACTGGCCATACCTGTTGAGAGGGCCA 739
785 GCATCCTAGGCATAAAGCTTGTATGACTAGAGAGTTGGCGCGAGGCCACCACTAGCG 844
740 GAGTTCTCGCGGTGAAGCTTTCGCGTGGGAGAG--GAGCGTGAACCCAGCGCGCT 796
845 TACATGGGCACGCTCTCTGCTCCCTGCGAGGCTTAAAGTAGATCTGTACGACTATGCGGAG 904
797 TGTACGGGCACCTACTCCGCTCTCTGGCAGGCTCAACGTAGACCTCTCTAGACTTTGCGGAG 856
905 AGATGCCAGATCAAGATCAAGATCTCGAGAGGTCGACAGATCTAGGCGGTGATCA 964
857 AGCTCCACGAGGTTAAGGTTAAGACTCTCGAGGAGTTCGCGGACTACCTAGGGGTGCTGA 916
965 ABAAGAGTGAACCGGTTATCATCAATTTGTTGGGAGATTCACGACTATTGGGACGACCGA 1024
917 GG-----GAGTACTGGATGACCCGT 937
1025 AGAAGAGACCACTATTACTGCAATACGCGCGAGCATGATCGCGCTACTTACGGCTTAG 1084
938 CAAAGCGGAGATACCTCAGGAAGTACTTAAGGAGGATGTCAAGAACCACTATGGGCTGG 997
1085 CCAGAGAAGATATTGCGGTTTGTCTATCCAGTTGTCTGATCAAGGTCTCCCACTAGACC 1144
998 CAGAGAAGTTCTTACCTTCCGGGCGCAGCTCTCAAGTCTCAGGGCTCCCGCTGAGCC 1057
1145 AGTAGTGGCATGAGTGTGGCTTTCGACTTGAATGTTACCTGATACCGGCGGCTTTA 1204
1058 AGGTTATGGCGGCGACGGTAGGCTTCAGACTGGAGTGGAGGCTCATACCGGAGGCGGCTA 1117
1205 AGATGAAGAGCTTGTGCGAAACCGGTTGAGCGCCCGACAGAGAGACTTACCGTGGCGCTA 1264

```

Db 1118 AGCTCGCGAGCTGGTCCCAACAGGGTTGAGAGGCGAGGCGCAGATACGCTGGGGCCA 1177
Qy 1265 TAGTCTTTCAGCGCTTGAGAGGCGGTGCAAGAGATATAGCCGTACTCGACTCTTTCAGTCA 1324
Db 1178 TAGTACTCAGCCCAAGCCGGGGTTCACGAGGATATGCCGCTCGACTTCGCAAGCA 1237
Qy 1325 TGTACCCAAACATCATGATTAAGTACAATGTTGGTCTGACAGCTTGTGAGGCTGGTG 1384
Db 1238 TGTACCCGAATATCATGTAAGTATAACGTCGGCCCGACACTCTGTCGAGCCCTGGGG 1297
Qy 1385 ABAAGTGTGCGAGTGGTGTCTGGGAGGCCCGGAGGTCAGACAGCTTCGAGGT 1444
Db 1298 AGAGTATGAGAGGAGGATATACAGGCTCCGAGGTGGGCCACAGTTTCGGAAGA 1357
Qy 1445 GTCCGCCGCTTCTTCAAGACAGTCTTTCGAGAGGCTGTAGAGCTTCGTAAGCGTGTGC 1504
Db 1358 GCGCTCCGCGCTTCTTCAAGAAATACCTTCGAGAGGTTCTTAAGCTCGAGGAGGAGNTTA 1417
Qy 1505 GTGCTGAATGAAGATATCTCCCGATAGCCGAGATATCGACTGTTCGATGAAGGC 1564
Db 1418 GGAGCGAGATGAAGACATCCGCGCCGACAGCCCGGAGTCAAGCTCTTCGACGAGGCG 1477
Qy 1565 ABAAGGCTTCAAGGTTCTTGCBAACCTAGTTACGCTACATGGTGTGAGCGGCGCTA 1624
Db 1478 AGAAGGCGATAAAGCTCTTGGCAACCCGAGCTACGCTACATGGCTGGCCCTGCTGCCA 1537
Qy 1625 GGTGGTATTCAGGAGTGGCGAAGCTGTCTCAGGCTTGGGTTAGGACCTCATACGCA 1684
Db 1538 GGTGGTACTGTAGAGAGTGTGCGAGGCGGTTCAGCGTGGGTAGGAGCATAATAGGA 1597
Qy 1685 CCGCCATACATAGTCTGTAACCTAGGCTCAAGGTATCTACGCTGACGACAGATTCGC 1744
Db 1598 CTGCTAAGAGAGCTGGGAGCTAGGCTGGAGGTATATATAGGAGACAGGACAGCC 1657
Qy 1745 TCCTTCGACCTATGATCCGAGAGGTTGGAATTTTCATCAAAATTAAGAGGAGGC 1804
Db 1658 TCCTTCGCAAGACATCTGAGAGGTGGAAGCTGTATAGTTTGTTCGAGAGGAGC 1717
Qy 1805 TGGGTTTCGAATCAAGCTAGAGAGGTGTCAAAACCTTATTTTCACAGAGCTTAAGA 1864
Db 1718 TTGGGTTTCGACATTAAGTGTGACAAAGTGTACAGCGGTATTTCTCACCAGGSCAAGA 1777
Qy 1865 AGAGGTACCTGGGCTTCTCGAGAGCGAGCTATAGATATGTGGTTTCGAGGCTGTAC 1924
Db 1778 AGAGGTATGTGGGCTGACGCTGGAGCGCAAGATAGATGTAGTCTGGCTTCGAGAGCTCA 1837
Qy 1925 GTGCGATTGTGTGNACTCCGCAAGAGGTTTCAGACTAAGTTGTGCAAAATAGTATTGA 1984
Db 1838 GGGCGGACTGAGAGGCTGCGAAGGAGCGAGTTTCAGGTGCGAGAGATGCTCTTAA 1897
Qy 1985 AGACGAGTGGTGAACAAAGCTGTAGAGTACGTGAGGAAGATTTGGAAGAGTTGAGG 2044
Db 1898 AGACGAGTGGTGGACGAGCTGTGAGTACGTGAGGAATATATAGAGAGCTTAGGA 1957
Qy 2045 AGGCGAGGTTCCCATAGAGAGCTGTGATCTGAGAGACCTTAGTAAGGCTTTCAGG 2104
Db 1958 GGGCCGAGTGGACATGAGGAATCTGGTTCATTTGGAAGCTCTGACACGCGCTCCATCA 2017
Qy 2105 AGTACACAAAGGAGCCACACAGCTGCTGAGCGAAGAGGATGCTGTACAGAGCTACC 2164
Db 2018 TGTACGAGGCGAGGAGCCCACTGACCGGAGCGCTACTTTATGAGGAGGCTGGCATAA 2077
Qy 2165 GGTGAAGCCGAGGCGAAGATAGGGTATGTAAATAGTAAGGGTGGTGGCGGTATCAGTC 2224
Db 2078 AGGTGAGCCAGGGGCTAAGTACGCTACGTGGTGTGACCAAGGGGTGAGGCGCCCTCTATA 2137
Qy 2225 ABAAGCATGGCCATATCTATGTCAGAGGATCTTAGCCAGATAGAGCTGACCTACTATG 2284
Db 2138 CGAGGCGAAGGCTTATTTTATGCTAGCAGAGGAGGTTTACGCTGGAGTACTATGTGC 2197
Qy 2285 TTGACCAACCAATCATCCCGCTGCATTGAGATAATCTGGCTACTTTGGCATCACGAGA 2344
Db 2198 ACAAGCAGGTGTGCGCGGCTTCTGCGCATATCTTCCAGTAACTTCGAGATTAATTGAGA 2257

Qy 2345 AGAAGCTGAAGACAGTCCAACTGGCGAGAGACTCTCTTCGACTT 2390
Db 2258 AGAGGCTGAAGGAGGAGGAA---GGCAGACAGCTCTCTCGACTT 2300

RESULT 13

AX035158 LOCUS 2610 bp DNA linear PAT 15-NOV-2000
DEFINITION Sequence 3 from Patent WO053772.
ACCESSION AX035158
VERSION AX035158.1 GI:11190912

KEYWORDS Pyrobaculum islandicum

SOURCE Pyrobaculum islandicum

ORGANISM Pyrobaculum islandicum

REFERENCE 1 Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;

AUTHORS Thermoproteaceae; Pyrobaculum.

JOURNAL Antranikian, G., Frey, B., Kaehler, M. and Sobek, H.

TITLE Dna polymerase from pyrobaculum islandicum

ABSTRACT Patent: WO 0053772-A 3 14-SEP-2000;

ANTHROPOLITAN GARABED (DE); FREY BRUNO (DE); KAEHLER MARKUS (DE);

SOBEK HARALD (DE); ROCHE DIAGNOSTICS GMBH (DE)

FEATURES Location/Qualifiers

1..2610

/organism="Pyrobaculum islandicum"

/mol_type="genomic DNA"

/db_xref="taxon:2277"

BASE COUNT 724 a 482 c 717 g 687 t

ORIGIN

Query Match 24.0%; Score 580; DB 6; Length 2610;

Best Local Similarity 55.0%; Pred. No. 1.1e-149;

Matches 1328; Conservative 0; Mismatches 1025; Indels 63; Gaps 7;

Qy 7 GAAGTTGTATTACCGGTTTTCAGTCTAGTACGAGGTTTGTGTAAGAGGCTCAGGTA 66

Db 133 GAACTAAAGTTTGGCTCTCGACATATCATATGCGTAGTTGGTAGTTCTCGATT 192

Qy 67 ATCATATGGGTTATGCTGAGAACGGCGAGAGGAGTAGTCTCTCATTTGACAGGCTTTTCG 126

Db 193 AGGATTTTCGTTATTTTAAAGCAGCGCTGAGCGTTGTACTTATCATGATAGATCTTTTAA 252

Qy 127 CCATATTTTATGCGCTGCTTCCACCGGCGCGGATCTTAAGCAGGTGACCAAGTATT 186

Db 253 CGATATTTCTA--CGTCGATTGTGCGTTTGTGAACAGCTGCTTAAAGCTGCGTTGT 310

Qy 187 CGTGCATTGAGTAGGCGCAAGAGCCCGATTATAGTGTAGAGGATGACAAAGAGGATAC 246

Db 311 CGCGAGTTGCG-----CCATAGACGATGTACAAATAGTTGAAGGAGTTTC 357

Qy 247 TTCGAGAGGCTCGTAGGCTTTACGATTCGACCGCTGCTACCCGAGGCTGTAGGAG 306

Db 358 CTCGAAGGCTTAAGAGTTTGTGAAGTTTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 417

Qy 307 TATCGCGAATCTGTAAGAGAGCTTGTATGTTGAGGATGTTCTAGAGCGGATATACGC 366

Db 418 CTTAGAGAGCGCCCAATGTGATACCAAGGTTTCTGGCGTTTATGAGCAGATATTCGT 477

Qy 367 TTCGCTATGCGCTATCTCATAGATCAGATCTATTTCTTTTCACCTGGTACCGGTAGAG 426

Db 478 TTTTATCATGGATATATGATAGATGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 537

Qy 427 GCTGAGCCCTCGAGAACAGATGCGCTTCCGTTGCTGACCAAGGATATACCTGGTTAAGAC 486

Db 538 GTAGAGAGGGGGGTAGACTTTGCGCGGATTTCCACATAGCTGTTCT-----583

Qy 487 AGCCCGGAGCCACTTTTATGTTGAGGCTCTCGACCAACCAAGCTTCCCGATCTTAGGATA 546

Db 584 -----CTCAGTGGTACGGGATGAGAGGCTTTCCCGCTTTCCCTTCATTGAAAGTT 630

Qy 547 CTGCGCTTCGATATGAAGTTTATAGCAAGAGGCTGCGCGGTCCAGAGCGCGATCCT 606

REFERENCE

2 (bases 1 to 3500)
Kaelher, M. and Antranikian, G.
Direct Submission
Submitted (14-Oct-1999) Technical Microbiology, Technical
University Hamburg-Harburg, Denickestr. 15, Hamburg 21071, Germany

FEATURES

source

1. .3500

/organism="Pyrobaculum islandicum"

/mol_type="genomic DNA"

/strain="GEO 3"

/db_xref="taxon:2277"

550. .2907

/gene="polB3"

550. .2907

/gene="polB3"

/notes="DNA-directed; no inteins present"

/codon_start=1

/transl_table=11

/product="family B DNA polymerase"

/protein_id="AAP27815.1"

/db_xref="GI:6752664"

/translations="MELKVPDIDITVAVGSPVEIRIFGILSSGRVWLIDRFPKPYF
YDCAVCEPAALKTALSRVAPIDDDQIVVERFLGRSKFLKVIKIPEDVRKLRAAM
SIPRVSGVVEADRFVYRMIDMVGVCNNVAEVEGRGLGIPYVVSQYGRDEG
PFSLKMAPDIEVYNERGSPDIRDPVWMLAKTNDGHEEVFEAGKDDRGVRAFY
DFIRSDPDVIVGNSGDMFYLVERAKAVGPLKVDRLSNPPOQSVVGHWSILVRA
NVDLYNIVEFPPEIKLIDRVAEYFGVMKREVRVLIHGKHYEYKWDPNKRLKRY
VLDDVRSTADKLFLFLQSSVGLPDQVAAASVGNRVEMMLIRYAIRLGEVAP
EGNVAPEVGHAIKLEPKPMYEDVLDFSSMYPNIMKYNLSPTLYLSEPEEDPP
NRREYVEPKGIVLEPKPMYEDVLDFSSMYPNIMKYNLSPTLYLSEPEEDPP
LKMANNAYGLCWGARWYKREAVSVTAFARAILKVDTEQARRUGIVVVDGTDLSL
FVKHSGDKLIIKVBKYGIDIKVDKDYAKVLEFTEAKRYAGLDRGDIIVGRFV
RGWSLAKDQURVIEIILKSDIKVDKDYAKVLEFTEAKRYAGLDRGDIIVGRFV
LKELEDEYKAYPPHAAQILKRRHYEVGKGTIGYIVVGGSEKVSERADPFIYLLDDI
KKIDIDYIERQIIPALRIAEVIGVKESDLKTRMERSLLDFLS"

BASE COUNT 955 a 674 c 902 g 965 t 4 others

ORIGIN

Query Match 24.0%; Score 580; DB 1; Length 3500;
Best Local Similarity 55.0%; Pred. No. 1.1e-149;
Matches 1328; Conservative 0; Mismatches 1025; Indels 63; Gaps 7;
7 GAAGTTGATTCACGGTTTATAGCTTACGCTACGAGGTGTGTGTAAGAGCCCTCAGGTA 66
553 GAACATAAAGTTTGGCCCTCTCGACATACATACATATCCCGTAGTGTGTCTCTGAGATT 612
67 ATCATATGGGTATTCGTGAGAACGGCGAGAGGGTAGTCTCTCATATGACAGCTCTTTTCG 126
513 AGGATTTTCGGTATTTAAGCAGCGGTGAGCGTGTGTGTACTTATCGATAGATCTTTTAAG 672
127 CCATATCTTATGCGTCTTTCGACCGGGCCGATCTCTAAGCAGGTAGACACAGCTATT 186
673 CCGTATTTCCTA - CGTCGATTGTGCGGTTTGTGAACCCAGCTGCTCTAAGACTCGCTGT 730
187 CGTCGATTGAGTAGGCCAAGAGCCGATTATAGTGTAGAGGTGACACAGAGGAAGTAC 246
731 CGCGAGTTGCG - - - - - CCAATAGACAGATGACAAATAGTTGAAGAGGTTTC 777
247 TTCGGAGCGCTCGTAGGCTCTTACGTATTCGACCGCTGCTACCGAGGCTGTGTAGGAG 306
778 CTCGAAAGGTCTAAGAGTTTGTGAAGTTATGCTTAAATTCCTGAGGATGTCCAGAAA 837
307 TATCGCGAACTCGTAAGAAAGCGTTGATGCTGTGTAGGATGTCTAGAGCGCGATATACG 366
838 CTTAGAGAAAGCGCAATGTCATACCAAGGGTTTCTGGCGTTTATGAAGCAGATATTTCGT 897
367 TTCGCTATGCTCTATCATAGATCAGATCTATTCTTTTCACTGCTACCGTGTAGAG 426
898 TTTTACATGCGGATATATGATACATATGGGGTAGTGGCCCTGTAGTTGAATGTGGCAGAG 957
427 GCTGAGCCCTCTCGAGAACAGATGGCTTCCGCTGTCGACAAAGGTATACCTGTTAAGAGC 486
958 GTAGAGAGGGGGGTAGACTTGGCGGGATTCCCATATAGTTGTCT - - - - - 1003

487 AGGCCGAGCCACTTTATGGTGAAGCTCTCGCACCAACCAAGCTTCCGATCTTAGGATA 546
1004 - - - - - CTCAGTGTACGGGATGACGAAGGCTTTCCCTCTTATTGAAGATT 1050
547 CTCGCTTCGATATGAAGTTTATAGCAAGAGGGTGGCGGTCTCAGAGCGGATCCT 606
1051 ATGGCGTTTGATATCGAGGTTTATAACGAGCGGGCTCTCCCGATCTCATCAGAGATCCC 1110
607 GTAATAGTAGTAGCTGTGAAGACTGACGATGCGGATGAGGTGCTATTCAATTGACAGAGGC 666
1111 GTGGTTATGTAGCCATTAAGACAAACATGGGCAAGAGGTTTGTAGAGCTAGTGGG 1170
667 AAAGACGATCGAAAAACCGATACGCGAGTTTGTAGAGTACGTGAAGAGGTATGACCCGAC 726
1171 AAAGACGATAGGGGGTTGTGCGGCTTTTGTGGATTATCAGAGTTTATGACCCGAT 1230
727 ATAATAGTCGGTTATAACAAACAATCATTTTCGATTTGGGCTTATCTTTTGAAGCGGCCGC 786
1231 GTAATTGTGGCTATAATTCCACCGGCTTTGATTGGCTTATTATTAGTAGAGCGTGGGAAA 1290
787 ATCTTAGCCATAAAGCTTGTGACTAGAGAGTTGGCGCGGAGCCACCACCTAGCGTA 846
1291 GCGGTCGGCGTCCGCTCAAGT - - - TGAATAGCTGAGCAATFCCCTCAACAGAGGCTG 1347
847 CATGGCGACGCTCTCTGTCCCTGGCAGGCTTAAACGTAGATCTGTAGCACTATGCGGAAAG 906
1348 TATGGACACTGCTCTATCGTTGGTAGAGCAATGTAGATCTCTATAACATCGTGGAAAG 1407
907 ATGCCAGAGATCAAGATTAAGAGTCTCAGGAGGTGCGAGAGTATCTAGGCGGTGATGAG 966
1408 TTCCCGGAGATTAAAGCTAAAGACTCTTGAACCGTGTGTGTAATATTTTGGTGAATGAAA 1467
967 AAGAGTGAACGGTTTATCATCAATTTGGTGGGAGATTCCAGACTATTGGGAGCAACCCGAG 1026
1468 AGAAGAGAGAGGTTGTATCCAGGCCACAGATTTATGATATCTGGAAGACCCCAAT 1527
1027 AAGAGACCACTATTACTGCAATPACGCGCGCAGATGTCGCGCTTACTTACGGCTTAGCC 1086
1528 AAGAGACCTTACTTAAAGCGGTATGTTCTAGACGACGTACGATCCACTTAGGACTTGGC 1587
1087 GAGAGATATTCGCGTTCTATCCAGTTGCTGAGTACAGAGTCTCCACTAGACAG 1146
1588 GACAGCTCTTACCGTTTAAATACAACTGCTTCTGTATCTGGGCTACCGCTGGATCAG 1647
1147 GTAGGTGCGATGAGTGTGCGCTTTCGACTTGAATGCTACTGATACCGCGCGCTTTAAG 1206
1648 GTGGCAGCGCGAGCGTGGGCAACAGGGTAGAGTGGATCTTCTTAGGTACCGTACCCG 1707
1207 ATGAAAGAGCTTGTGCGAACCGCGTTGAGCGCCCAAGAGAGACTTTACCGTGGCGCTATA 1266
1708 CTGGCGAGGTAGCTCCGAAACAGAGAGAGAGGGAGTACGAAACATATAAGGGGGCCATC 1767
1267 GTTCTTGAAGCGGTGAGAGCGGTGACAGAGATATAGCGGTACTCCAGCTTACTCGATG 1326
1768 GTGCTTGAAGCAAGCCCGGATGTACGAAGAGCTGCTTGTACTTGACTTCTTCTTCAAG 1827
1327 TACCAACACATCATGATAAAGTACAATGTTGGTCTTGCACACGCTTGTGAGGCGCTGTGAA 1386
1828 TACCTTAACATCATGATGAAGTACACCTATCCCGGATACGTACCTAGAGCCCGCGAG 1887
1387 AAGTGTGGCGAGTGTGCTGGGAGGCGCCCGGAGGTCAAGCACAGGTTTCGTAGGTGT 1446
1888 CCGGATCCGCGCGAGGTTGTAACGTTGGCGCCCGGAGTGGGCGATAGGTTTAGGAGNAGC 1947
1447 CCGCGCGGCTCTTCAAGACAGTTCTTGAGAGGCTGTAGAGCTTGTAGAGCGGTGCGT 1506
1948 CCACCAAGGCTTCGTCCTTCAAGAGCTTGTGAAGAGCTTGTGAGAGCTTGAAGAGCGGTAAG 2007
1507 GCTGAATGAGAGATATCTCCGATAGCCAGAGATATCGACTGTGTGATGAAGAGAGCAG 1566
2008 GAGGAGCGAAGAAATACCCCGGAGATTCCCGAGAGTTTGAAGATCTTAGACGAGAGACAA 2067

1567 AAGCGTTGAGGTTCTTGCAACAGCTAGTTACCGCTACATGGTTGGACCGCGCTAGG 1626
D5 CGTCCCTCAAGGTTATGCGCAATGCCATTATGATGATCTGGCTGGTGGCGCGCG 2127
QY TGGTATTGCGAGGAGTGGCAAGGCTGTCACGGCTGGGTTAGGACCTCATACGAC 1686
D5 TGGTATTGCGAGGAGTGGCAAGGCTGTCACGGCTGGGTTAGGACCTCATACGAC 2187
QY GCCATCAACATAGCTCGTAACTAGGCTCAAGGTGATCTACGGTGCACAGATTCGCTC 1746
D5 GTTATTGACCAAGCTAGAGGTTGCGCATTTGTTGTTATGCGATACAGACGCTA 2247
QY TTCTGTCCTATGATCCGAGAGGTTGGAATTTTCATCAAAATATATAAGAGGAGCTG 1806
D5 TTCTGTTT---CAAAACATGAGAGCTGCGACAACTGATCAAGTACGTTGAGGAGGATC 2304
QY GGGTTGGAATCAAGCTAGAGAGGTTGACAAAGCTTATCTTTACAGAGGCTAAGAG 1866
D5 GGCATAGACATAAAGGTTGACAAAGCTTATCTTTACAGAGGCTAAGAG 2364
QY AGGTACGCTGCTCTTCGAGAGCGGATGATATTTGCGGTTTCGAGGCTGTAGCT 1926
D5 AGGTACGCTGCTCTTCGAGAGCGGATGATATTTGCGGTTTCGAGGCTGTAGCT 2424
QY GCGGATTTGTTGAACTCCCAAGAGGTTGAGACTAGGTTGTCGAAATAGTATTGAG 1986
D5 GGAGACTGGAGTGAATTTGCCAAGAGCTTCAGCTAAGAGTTATAGAGATTATCTTAAG 2484
QY ACAGTGAAGTTGACAAAGC-----TGTAGAGTACGTCAGGAGGATGTGAA 2034
D5 TCAAGAGATATTTGCGAAGCTAGACATGGCGTTTAAATATATATAGAGAAATATAGAA 2544
QY GAGTTGGAGGAGGCGCAAGGTTCCCATAGAGAGCTTGTAACTTGGAGACCTTTAGTAAG 2094
D5 CGACTGAAATATTAAGTTTAACTTGAATTTAAATATGAAATCTAGACAAA 2604
QY GCTCTTGGAGGATACAAAGCGAGGACACACGCTGTTGCGAGGAGAGATGCTGTCA 2154
D5 GAGTTAGACGAATATAAGCGGTATCCGCTCATGTTTCAGCTGCGCAAAATTTCTCAAGAGA 2664
QY CGAGCTACCGGTTAGCCCGGCGACAGATGATGTAATAGTGAAGGTTGGTGGC 2214
D5 CACGCTATCGAGTGGCGGAAAGCGCTAGATGAGCTATGCTATGTTTAAAGGAGCGAA 2724
QY CGTATCAGTCAAGAGGATGCGCATCTTATGCTCAAGGATCTTACGATGAGACGCTG 2274
D5 AAGATATCTGAGAGGCGTTGCGGTATATCTCTTACGATATATAAAGAGATGATATC 2784
QY ACTTATGTTGACCAACCAATATCTCCGCTGATGAGATGATGAGTACTGGCTACTTGGC 2334
D5 GACTACTACATAGAGAGACAGATAATACCGCGGCTTGAAGATAGCAGAGGTTATAGGG 2844
QY ATCCCGGAGAGAGCTGAACCAAGTGCACCTGGCGAGAGACTCTCTTCCACTTCTA 2394
D5 GTGAAGAGTGGATCTTAAACGGGAGAA---TGAAGAGTCCGCTTCTGATTTCTA 2901
QY GCCAAGAGAGCAAGT 2410
D5 AGTTAGAAACTACGT 2917
RESULT 15
AP000988/c 266050 bp DNA linear BCT 25-DEC-2002
LOCUS Sulfolobus tokodaii genomic DNA, complete sequence, section:8/10.
DEFINITION AP000988 BA000023
ACCESSION AP000988.1 GI:15622956
VERSION
KEYWORDS
SOURCE Sulfolobus tokodaii
ORGANISM Sulfolobus tokodaii
ARCHAEA: Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
Sulfolobus.
REFERENCE 1

AUTHORS

Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Anka, A., Kobayashi, H., Hoshino, A., Fukui, S., Nagai, Y., Nishijima, K., Otsuka, R., Nakazawa, H., Takamiya, M., Kato, Y., Yoshizawa, T., Tanaka, T., Kudo, Y., Yamazaki, J., Kushi, N., Ouchi, A., Aoki, K., Masuda, S., Yanagii, M., Nishimura, M., Yamagishi, A., Oshima, T. and Kikuchi, H.
Complete genome sequence of an aerobic thermophilic crenarchaeon, Sulfolobus tokodaii strain7
DNA Res. 8 (4), 123-140 (2001)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

gene

CDS

gene

CDS

gene

CDS

1. 266050
/organism="Sulfolobus tokodaii"
/mol_type="genomic DNA"
/strain="7"
/db_xref="taxon:111955"
288..830
/gene="ST1855"
288..830
/gene="ST1855"
/note="OWL:ADOXF24H1 percent identity: 83.432 in 169aa."
/codon_start=1
/transl_table=11
/product="180aa long hypothetical protein"
/protein_id="BAB66946.1"
/db_xref="GI:15622957"
/translation="MSKINPNEGNAFLEVIPIFAVGMFLDGLKAVLNPAKLD
PNSSPVGKLVNPLPHAGFPKFLMTLENHAPDTFLTIFSYIAGLFLVIGLL
TRLAAGAAVNAIGMAPYMWIGTCEDEWQIGSLLTAGAIVVMTGTAGRVNGLDIFLY
KTFGDRGIANVPILKWLW"
824..1342
/gene="ST1856"
824..1342
/gene="ST1856"
/note="OWL:ADOXF24H2 percent identity: 68.590 in 156aa."
/codon_start=1
/transl_table=11
/product="172aa long hypothetical protein"
/protein_id="BAB66947.1"
/db_xref="GI:15622958"
/translation="MVKRMNKVAIVGIIFALFTVWILGTQWYGVNVPFNNSKL
PKLNTVITAYNTFQGLCLIMTIDIDPDYAPASAPLMEISNGTWHIFLNSQIAND
TVKLIQAPWNANKKDSVNWYSGFVILGSEQFHLAIKGLHLSFGKYEVKLYTPAIS
NRQAIAYETISS"
1378..2772
/gene="ST1857"
1378..2772
/gene="ST1857"
/note="PIR:D64486 percent identity: 34.286 in 105aa."
/codon_start=1

```

/transl_table=11
/product="464aa long conserved hypothetical protein"
/protein_id="BAB66948.1"
/db_xref="GI:15622959"
/transl_table=11
/product="MENSKERIRASQIYSDGILGLKRGKLIIGVKTIDIVRVE
YDLSLKGCECELCNGCHIVAIKMSYEGVIDPDSLENIENKRELIGILVTL
IEKPPMIANYIPIENAKVSLERYINLIQNGENIVNSFTPLINRKRKINDIFI
LDITASCKSPYNTITEKPDENIMKTLANILLEKEVEDDIREKLEIKIEKQIGN
LDFTVLTDNEDIRKIMDIRIYNALIRGDKDKILKLLQTDVSKKEFNILLQTD
EKALFAKMLSSLFNYYNLGFSQALNKKMIELKDIIGISSHKRILPLIK
GNPDLKSLVLSKDNWVLPILINLYDVASGLKYDIAVTWMDKFSKQCPDVR
IUCQEKELSVIQLHTEELVERKEYEDVIQCLVKARKYMTIEDFNLLSQIKENYK
RKQLVSLNKILS"
2952..4544
/gene="ST1858"
CDS
2952..4544
/notes="ST1858"
/notes="PIR:B69172 percent identity: 24.545 in 110aa."
/codon_start=1
/transl_table=11
/product="530aa long hypothetical protein"
/protein_id="BAB66949.1"
/db_xref="GI:15622960"
/transl_table=11
/product="MSQEKSLFRESSGLRVSPWASMSATFGLVGTGVPILILSW
LTPAGNWTALFLLPPTLWAFLEYVAGISMPRAGGDYVENSRAVGFVNVW
GLTFGALSGLYSYLAQWFAVLFSGIGLAYSNQYFLNIGKELGTBAVILGTIV
IISTALAMIPRQKQVLMGGIISLSSIIIMFIALAQISPSQALASRVGPNAYS
EVIONAESGLTFPALYASFLALPVVYTYWYNTPASGEMKPRKNVLSIILA
ILVTSAYIILVIFSEKAFASLTYSYNGINDVYNTLSNIGDTPYPAFTN
HSLPRTIMEILWLPNFYSPPLVALVRYLFWGDFRVPMPPLADVDRFKPSLKA
TVLVLGELGVLVLYAFNTPVAIVDTTVPEVGVAFISTALMPFRVLDLYKRAVQH
KNINAGIPVISIGFVGLFMVHLITWNNPVLIPNFVTLISLGIYAGVLYIYI
GYLLNRRRGIDLLLFNEVPPE"
4578..6440
/gene="ST1859"
CDS
4578..6440
/gene="ST1859"
/notes="Swiss_Prot:Q46819 percent identity: 32.759 in 58aa.
motif:prokaryotic membrane lipoprotein lipid attachment
site"
/codon_start=1
/transl_table=11
/product="620aa long hypothetical protein"
/protein_id="BAB66950.1"
/db_xref="GI:15622961"
/transl_table=11
/product="MNLILFSAISMLLVFVYVYVKNWSDRRVLYVAYVEGMM
VMNLGAYLLHDLFLFLVINGAYMIFGLYALINIKSISRYAYVLFSSFMISEV
MGSLFYIOTGLPANDLSSVENPFWTVWAEIMFTLWVSFKRLDKTLRNLIISLVL
MPFPVIFSNVAPLFSIIMIGATILAVETLYNQLRATQOTFTTIELMSIFLMMIG
GTYLFLFSLTVDVSMLIANVFIYTVIGPAPRKONYLRTDKIAFSIIFLPEIMEF
FMGVLDPFVWGLFSPGKGFSSLSLPTWNIWDGIDVGSILGSPWFLIMGIEWG
FLAPKMLERVREVRKMLMILAYAIYIYPSFPLSSSLFPIYPMWSMGITLA
PVNSVLIGLIGTYIYVAILSFGLSRLNCAVTCAPMYQGNFYDSLKYVNRSTIG
KLLTSSKSYRPTALSVLVILAAVSYLNSLGIISPTIFNDIVLYIYTFWFDIM
WYLIISIPLGTACTVTCYVGVNFQALSRIGLFRLVKQNPQOICECTYVDCANA
CPVGLTDNRGEFIRKGRFSIKCVGICECIDACYPNNIPIYDVNRNWERFRGK"
6481..7209
/gene="ST1860"
CDS
6481..7209
/gene="ST1860"
/codon_start=1
/transl_table=11
/product="242aa long hypothetical protein"
/protein_id="BAB66951.1"
/db_xref="GI:15622962"
/transl_table=11
/product="MWVAILVYTVLYSVSYKIKTWIGLFFAPFVLLIMMTWIFGIAVYL
LSPNISLABAIIYVNASMLILLYLPMNAKLSNDESKIHIFSLILTVLNEILM
GGAFSLAFGKIFELTYFSAFLTILNSWFFYPMNAEMALYILIDYKSRANKELFAL
IGITFFPTPIYFNSQWYSSIFIDVILSLIGIINSGRLWRYLYIITAISSVVVDFIPT
FFDIIIIIDMILYVYLLNRKVS"
complement(7183..8685)
/gene="ST1861"
CDS
complement(7183..8685)

```

```

/gene="ST1861"
/notes="Swiss_Prot:P26235 percent identity: 25.946 in
370aa."
/codon_start=1
/transl_table=11
/product="500aa long hypothetical Na(+)/H(+) antiporter"
/protein_id="BAB66952.1"
/db_xref="GI:15622963"
/transl_table=11
/product="MSLLVDSLLYLIMLILAKAEAFSLNLVRFVGPFIIVGVILG
SGVLDLVKINIIISFITSLSGLVFLFAGAEELGKLVESKIDFSLIELSLPILII
SILLYFKIFLIVPLAMTSAGFLRLDLGLSKHKLALFYQSVLNEIVAVVL
PAIFSKFOILDIISILLVFLFGLGGKIAKLEIEGVYKVRIEFATISVFIIG
YVAEVSNSAITALFLGLRDLKORPYLIERLAFTYGFPEFLFVSGLYFVKI
SPTLILSLTELATIGSKFAGLSAKIQIDVINGLGTSMKGVDVSLIISALTS
MLTPYQXSFLSLAITRSSLVLPFLFMKFKPKIOENKINLPVLNLIKQVTVNC
YQTLREAIGIINERGYRAIVVDFRPLGYVSVQSLLEIDPNDTEIDKVCDDIYNEV
KILEKKSXVIDLVRFRETEPVPVVIVDDGRLLSVVYRELLRYLTPIQ"
8806..9507
/gene="ST1862"
CDS
8806..9507
/gene="ST1862"
/codon_start=1
/transl_table=11
/product="233aa long hypothetical protein"
/protein_id="BAB66953.1"
/db_xref="GI:15622964"
/transl_table=11
/product="MELEYNEFTNTPKDVILEYISNPNELKYVPAPKSLKRLDENS
W
BLEVKMLTILKVKRIITGDEVYTIETKGLIKISLSLKYVILTPKRNITLIKIF
SYEGPFESIAKQTEEXYRRGVEIFRDEKLEKLNQIRVTKTFDLNWKTLIAKG
IKRGEDILFRAMIESYRNLVLSDDKTIVELIFENSLSDSKGDIINTLGENIKY
LMKISQIMETKNRTS"
complement(10064..10783)
/gene="ST1863"
CDS
complement(10064..10783)
/gene="ST1863"
/codon_start=1
/transl_table=11
/product="239aa long hypothetical protein"
/protein_id="BAB66954.1"
/db_xref="GI:15622965"
/transl_table=11
/product="MSLPMVNVNRYIYFHOEPMSKISVVKPLVENDQITITPSI
LPLITNIIYFVNFIIISLSEKIFVILISILPLSYLGLSKVNSKTIYIYVLSIL
FINTKIGLVTLISILLIPIYFTHRSKYAFASATANLIYEFVSKDSYISLVKSDL
RLSPYKMYLELISNNSGILFHDFTIYALNKGVRGIIIEVEYVYKDMRKNKPELLNT

```

```

Query Match 23.8%; Score 574.6; DB 1; Length 266050;
Best Local Similarity 54.9%; Pred. No. 5.1e-148;
Matches 1288; Conservative 0; Mismatches 1014; Indels 45; Gaps 6;

Qy 16 TTCACGGTTTGTAGCTCTAGCTACGAGGTGTTGGTAAAGAGCCTCAGGTAATCATATGG 75
Db 216874 TTCTTATCTTGTGACTCTCATATGATGATGTAGTTGAAATAAACCCTTATTTATATTGG 216875

Qy 76 GGTATTGCTGAGAACCGGAGAGGGTGCCTCATTCAGCAGGTCTTTTCGCCCATCTTC 135
Db 216814 GCTATTGATAAAGAGGGTAATAGGGTTGATTATTAGAGAAAAAATTCGCTCCATATTC 216755

Qy 136 TATGCGCTGCTTGCACCGGGCCGATCCTAACGAGTAGCAGCAAGTATTCGTGCATTG 195
Db 216754 TATGCGTTGCTAGATGATAACTATATATATTGTAATTTAGAAAAGAGATATTAACACTT 216695

Qy 196 AGTAGGCCAAAGACCGCGATTATAGGTGTAGAGGATGACAAGAGGAGTACTTCGGGAGG 255
Db 216694 AGTAAACCATCTCGCGCGATTACTTCAATAAAGAGTAGAAGAAAATCTTTGGCTCA 216635

Qy 256 CCTCGTAGGCTCTTACGTATTCGACCGGTCTACCCGAGGCTGTGTAGGAGATATCCGAA 315
Db 216634 CCTGTAAAGGTTTTTAAATAGAAACAGTAATTTCCAGCTTACGTAAGATATATAGAT 216575

Qy 316 CTGCTAAGAACCTTGATGTGTGTAGGATGTTCTAGAGCGCGATATACGCTTCGCTATG 375
Db 216574 GAAGTAGCAAAAATAAAGAGGTAAAGAGTGTGTTTAGAGCGCATATTCGCTTTTACATG 216515

```


QY 376 CGCTATCTCATAGATACAGATCTATTTCCCTTACCTGGTACCGGTAGAGGCTGAGCCCC 435
Db 216514 AGATATTCGATAGATATAATCTAAAGCCTTCTATTGGATAGAGGCTGAAGTGAAGAA 216455
QY 436 CTCGAGACACAGATGGGCTCCGTGTCGACAGGATACCTGGTTAGAGCAGCCCGAG 495
Db 216454 ATAAAGAGAGATAAT---TTTAGAGTCAAAAGAGTTTATGAATTAAGAAATAATAAG 216398
QY 496 CCACCTTATGCTGAGGCTCTCGCACCAACCAAGCTTCCCGATCTTAGGATACCTCGCGTTC 555
Db 216397 CTTTATGAGAT-----AAATACAGAGCTTAAAGTTTGGCCCTT 216356
QY 556 GATATGAGTTTATAGCAAGAGGTCGCGGTCGAGAGCGGATCCCTGTAATAGTG 615
Db 216355 GATATGAGTTTATATAAATATGTTCTCCCAATCCAGGAGATCTCTGTTATAATA 216296
QY 616 ATAGCTGTAAGATGACAGATGGGATGAGGTCTATTTCATTCGAGAGGCAAGACGAT 675
Db 216295 ATTGAGGTATGACCAAAAGAGGAGTAA---ACATTTTATGCTGATAGTATGATGAT 216239
QY 676 CGAAACCGATACCGGAGTTTGTAGATGACGTGAAGAGGTATGACCCCGACATATAGTC 735
Db 216238 TTAAAGGCTATACGTGAATTTATTAATTTGTACAAACCTATGATCTCGATTTGTA 216179
QY 736 GGTATACAAACATCTTCGATTTGGCTTATCTTTTGGAGCGCGCCGATCCTAGGC 795
Db 216178 GGATATAATAAACAATTTGATTTGGCCCTATTATTAGAAAGGCGCAACATAGAGGA 216119
QY 796 ATAAAGCTTGTATGACTAGAGAGTTCGGCCGAGCCACACCTAGCGTACATGGCAC 855
Db 216118 ATTAGATTAGATGTAGGAGAGAGTTAATGCTGAACCTCTCAAGGATTTATGGACAT 216059
QY 856 GTCTCTGCTCCGCGAGCTTAACGTAGATCTGACGACTATGCGGAGAGATGCCAGAG 915
Db 216058 TACTCGATTACTGGGAGTTAAATGTAAGTATATGTTTGGCGACTCTACAAAGAA 215999
QY 916 ATCAAGATAAGAGTCTCGAGAGGTCGACAGATCTAGCGGTGATGAAGAGAGTGAA 975
Db 215998 GTAAAGATTAAACTCTCGAAATATAGCAGATATTATGAGTGTATTACCTTAAGGAGAA 215939
QY 976 CGCGTTATCATCAATTTGTGGAGATTCCAGACTATTGGGACGACCCGAGAGACCA 1035
Db 215938 AGAACAAATGTGAATGTATGATATTCGAGATTTGGGATGTAGAGAGACGAGAT 215879
QY 1036 CTATTACTGCAATACGCGCGAGATGTCGCGCTACTTACGCTTACGCGGAGAGATA 1095
Db 215878 ATATTACTTAATAACAATCTAGATGATGCAAAATCTGCGTATCTTTTAGGAGAGTATTC 215819
QY 1096 TTGCGGTTGCTATCCAGTTGTCGTACGTAACAGTCTCCACTAGACCGGTAGTGCG 1155
Db 215818 ATACCTTTGGGATTTGAATTAAGATTTAGCGGATTAACCACTAGATCAACTTCCATG 215759
QY 1156 ATGAGTGTGGCTTTCCGACTTGAATGTTACCTGATACGCGGGCTTTAAGATGAAGAG 1215
Db 215758 GCTAGTGTGGCCATAGAGTTGAGTGTATTAATGAGAGGAGATATAATATATGA 215699
QY 1216 CTTGCGCGAACCGGTTGAGCGCCGAGAGAGACTTACCGTGGCGCTATAGTTCTTGAG 1275
Db 215698 CTTATTCCAAATAAAGAGAGAGAGAAATAGAGGTATGAGGCGGTTTAGTAATCTCA 215639
QY 1276 CCGTTGAGAGGCGTCGACAGAAATATAGCCGTACTCGACTTTAGCTCGATGTACCAAC 1335
Db 215638 CCATTACCGGGCATCATGAAGATGTTTATGTACTAGATTTCTCGTCAATGTATCCCTCA 215579
QY 1336 ATCATGATAAGTACAAATGTTGCTCTGACAGCTTTGTGAGCCCTGGTGAAGAGTGGC 1395
Db 215578 ATTTGATTAATAATATACTTTGCTCCAGATACGCTGGTAAG-----GGA 215534
QY 1396 GAGTGTGTTGCTGGGAGCCCGGAGGTCTAGCACAGGTTCCGTAGGTGTCCGCCCGGC 1455
Db 215533 GAATGTGAATAATTTGTTGATCTCTCCAGCGGGCATAAATTTAGAAAGATCTCCAGGA 215474
QY 1456 TTCTTCAAGACAGTCTTGAGAGGCTGTAGAGCTGTGAAGCTGTCCGTGCTGAATG 1515

Db 215473 TTATATAAAATGTACTTGAATAAACTAATACAGAGAGAAAGGTAAAGATTAATG 215414
QY 1516 AAGAGTATCTCCCGATAGCCAGAAATATCACTGTTTGGATGAAGGAGAGGCGGTTG 1575
Db 215413 GAAAAA---ACATGGATGAGTATGATAAGAGGTGTAGATCGAGACAAAGACTCTA 215357
QY 1576 AAGTTCTTTGAAAACGCTAGTTACGCTACATGGGTTGGAGCGCGCTAGGTGTTGTC 1635
Db 215356 AAGTAATGCAAAATGCCCTTTTATGTTATATGGGTTGGTGGTGAAGATGGTACAGT 215297
QY 1636 AGGAGTGCAGAAAGCTCTGACGCTTGGGTAGGACCTCATACGACCCGATCAAC 1695
Db 215296 AAGAGGGTGCAGAGCTCTTACGCTTCGGGAGACAAATAATCTGAGTCAGCAAAA 215237
QY 1696 ATAGCTCGTAACCTAGCGCTCAAGGTGATCTACGGTGACACAGATTGCTCTTCGTACC 1755
Db 215236 ATTGCTAAAGAAAAAGGTTTACGTAATATATGTTGATCTGATTTTGTTA-- 215179
QY 1756 TATGATCCGAGAGAGGTGGAATTTTCATCAAAATTTATAAGAGAGAGCTGGGTCGAA 1815
Db 215178 -AGGAGGAGGGGATATTAAATTAATTAACAGAAATATCTTCCAAAGTTCCGGGTAGAA 215120
QY 1816 ATCAAGCTAGAGAGGTGTACAAAGCTTATTCTTTACAGAGGCTTAAAGAGGTACCT 1875
Db 215119 ATTAAATATGATAAAATATATAAAGAGTATTCTTTACTGGAATAAAGAAACGCTATGCT 215060
QY 1876 GGCTTCTCGAGACGAGATGATAGATATTGCGGTTTCGAGGCTGTACGTGCGCATGG 1935
Db 215059 GGATTAATGAGATGCTGAAGATAGATATTGTTGGTTTGAAGCTGTAGAGAGACTGG 215000
QY 1936 TGTGAATCCGCAAGAGGTTCAGACTAAGGTTGCGAAATAGTATTGAAGACGAGTAG 1995
Db 214999 TGTGATTTAGCTAAGCAAGTTCAAACTAATGTGATGAGTAAATCTCAATCTGGAAG 214940
QY 1996 GTCAACAGGCTGTAGAGTACGTCAGAGAGATTGTGAAGAGTTGAGAGGCGCAGGTT 2055
Db 214939 GTTGAGGTGCAATAAATAACGTTAAACAGTCATATTTGATTTAAGAGGTATAACTTT 214880
QY 2056 CCATAGAGAGCTTCTTAATCTGGAAGACCTTTAGTAAGCGCTTTCGAGAGTACACAG 2115
Db 214879 AGATAGAGGATTTAATAATATGAGAGACGATCGATAAATACTTATGATGTT 214820
QY 2116 GAGGACCAACAGCTGCTGAGGAGAGAGATGCTGTCAGAGGCTTACCGGTTAAGCCCA 2175
Db 214819 ACAGCCCTCATGTAGTTGAGCTAAAAAGCGCAAGGCTGATATTAGTGTCTAAA 214760
QY 2176 GCGGACAGATAGGTTGATGTAAGTGAAGGTTGGCGGTATCAGTCAAGAGCATGG 2235
Db 214759 GGTGTTAAGATAGTTACGTTATTGTAAGGGATCTGGAAAAATTTCTGATTAAGCTGAA 214700
QY 2236 CCATCTTCTATGCTCAAGGATCTTAGCCAGATAGAGTCACTTATGTTGACCCCA 2295
Db 214699 CCATCTTCTGATTAAGAGAAAAACAGATTGACGTAGATATTATATAGATAACAA 214640
QY 2296 ATCATCCCGCTGCATGAGATPACTGGGCTACTTTGGCATCCCGAGAGAGAGCTGAA 2355
Db 214639 ATAATACCACTTGCATGAGATTTTGAAGGGTTTGGAGTAAAGAAAGTTTCACTTAA 214580
QY 2356 GCAAGTG 2362
Db 214579 ACTGGTG 214573

Search completed: November 26, 2003, 16:16:16
Job time : 8735 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 26, 2003, 11:55:14 ; Search time 4961 Seconds
(without alignments)
11816.649 Million cell updates/sec

Title: US-10-034-849-1
Perfect score: 2412
Sequence: 1 agcactgaagtgtattcac.....tagccaagaagcaagtaa 2412

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estmu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_atci.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_htc.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_estom.*
- 17: em_gss_hum.*
- 18: em_gss_inv.*
- 19: em_gss_pln.*
- 20: em_gss_vrt.*
- 21: em_gss_fun.*
- 22: em_gss_mam.*
- 23: em_gss_mus.*
- 24: em_gss_pro.*
- 25: em_gss_rod.*
- 26: em_gss_phg.*
- 27: em_gss_vrl.*
- 28: gb_gss1.*
- 29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59.4	2.5	706	9	AI557857 P6test.G1
2	57.4	2.4	579	28	AZ047965 LNAJFV1.1
3	57.4	2.4	1053	29	AL392960 T7 end of
4	54	2.2	483	14	C16083 C16083 Clon

5	54	2.2	659	9	AL892989
6	52.4	2.2	576	9	AV988988
7	52.4	2.2	595	10	BG656636
8	52	2.2	389	14	C16480
9	51.8	2.1	352	9	AI211172
10	51.8	2.1	379	9	AI211173
11	50.6	2.1	425	28	AZ047964
12	50.4	2.1	674	14	CD283171
13	50.4	2.1	710	14	CA854286
14	50.4	2.1	752	12	BI889706
15	49	2.0	922	10	BF136910
16	48.6	2.0	462	14	C16462
17	48.4	2.0	1201	13	EX381961
18	46.4	1.9	636	14	CB176554
19	45.8	1.9	585	29	BZ293548
20	45.2	1.9	599	10	BG738046
21	44.2	1.8	885	13	EX425603
22	43.8	1.8	890	13	EX075042
23	43	1.8	553	12	BI674029
24	42.4	1.8	680	14	CS012480
25	42.2	1.7	503	10	BE535321
26	42.2	1.7	568	10	BF203965
27	42.2	1.7	700	12	BM008621
28	42.2	1.7	706	10	BF346914
29	42.2	1.7	724	12	BG826841
30	42.2	1.7	739	10	BG744446
31	42.2	1.7	758	10	BE547846
32	42.2	1.7	788	10	BF205093
33	42.2	1.7	810	12	BM008549
34	42.2	1.7	888	14	CA455120
35	42.2	1.7	928	13	EX350425
36	42.2	1.7	953	10	BF304095
37	42.2	1.7	1013	13	EX366475
38	42.2	1.7	1020	13	BQ070749
39	42.2	1.7	1028	12	BM479873
40	42.2	1.7	1056	10	BG120642
41	42.2	1.7	1067	12	BQ054258
42	42.2	1.7	1201	13	EX381961
43	42.2	1.7	1201	13	EX382861
44	42	1.7	669	13	BU060437
45	41.8	1.7	760	9	AUI21118

ALIGNMENTS

RESULT 1
AI557857
LOCUS
DEFINITION P6test.G1.r misc Homo sapiens cDNA 5', mRNA sequence.
ACCESSION AI557857
VERSION AI557857.1 GI:4490220
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 706)
AUTHORS Huang,G.M., Ng,W.L., Farkas,J., He,L., Liang,H.A., Gordon,D., Yu,J. and Hood,L.
TITLE Prostate cancer expression profiling by cDNA sequencing analysis
JOURNAL Genomics 59 (2), 178-186 (1999)
MEDLINE 99339982
PUBMED 10409429
COMMENT Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.
Location/Qualifiers

FEATURES

```

source
1. .706
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="misc"
/note="Organ: Prostate; Vector: pBluescript; Directional
cDNA library was constructed using Lambda Zp II kit
(Stratagene). mRNA was extracted from a prostate tissue."
BASE COUNT 192 a 119 c 154 g 185 t 56 others
ORIGIN
Query Match 2.5%; Score 59.4; DB 9; Length 706;
Best Local Similarity 49.7%; Pred. No. 1.7e-05;
Matches 144; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
QY 1805 TCGGGTTCGAATCAAGCTAGAGAGGTGTACAAAGCTTATCTTTTACAGAGGCTAAGA 1864
Db 37 TGGNCTGCTAGAGCTTGAATATGAAAGGGTTTATATCGAGGGGATCTCTCGTTACGAAGA 96
QY 1865 AGAGGTACGCTGGCTTCTCGAGGACGAGCTATAGATATCTCGTTTCGAGGCTGTAC 1924
Db 97 AGAGGTATCGATATAGATGAGAGGAGGAAAGTCATTACTCGTGTTCGATAGATAGTA 156
QY 1925 GTGGGATTTGGTGAATTCGCAAGAGGTTTCAGACTAAGTTTCGAAATAGTATTGA 1984
Db 157 GCGAGATTGGAGTGAAATTCNCAAGAACTCAAGCTAGAGTTTTCGAGACAATACTAA 216
QY 1985 AGACGAGTAGGTGACAAAGGCTGTAGTAGTACGTGAGGAAGTATGTGAAAGTTGGAGG 2044
Db 217 AACCGGAGATGTTGAAGAAGCTGTGAGAATAGTAAAGAAAGTAATACAAANCTTGCCA 276
QY 2045 AGGCGAAGTTCCTCAGAGAGGCTGTATCTCGAGACCTTAGTAG 2094
Db 277 ATTATGAATTCACAGAGAAGCTCGCAATTTATGACGATTACANG 326

```

```

RESULT 2
AZ047965
LOCUS
DEFINITION
LMAUFV1_lm71d05.x1 Leishmania major FV1 random genomic library
Leishmania major genomic clone LMAUFV1_lm71d05 3' similar to
SW:DP0D SOYEN 048901 DNA POLYMERASE DELTA CATALYTIC CHAIN 1,
genomic survey sequence.
ACCESSION
AZ047965
VERSION
AZ047965.1 GI:7257911
KEYWORDS
GSS.
SOURCE
Leishmania major
Leishmania major
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE
1 (bases 1 to 579)
Akopyants, N.S., Clifton, S.W., Martin, J., Pape, D., Wylie, T., Li, L.,
Kissinger, J.C., Roos, D.S. and Beverley, S.M.
A survey of the Leishmania major Friedlin strain VI genome by
shotgun sequencing: a resource for DNA microarrays and expression
profiling
Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
JOURNAL
MEDLINE
PUBMED
COMMENT
11295190
Other GSSs: lm71d05.y1
Contact: Akopyants, NS / Beverley, SM
WashU Leishmania Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Library construction: Natalia S. Akopyants, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
If using this information please cite:
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
Friedlin strain VI genome by shotgun sequencing' and the Washington
University Genome Sequencing Center For information on obtaining

```

```

clone material please contact: Natalia S. Akopyants Ph.D.
(natalie@boxcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
(beverley@boxcim.wustl.edu)
Seq primer: -40UP from Gibco
Class: shotgun
High quality sequence stop: 342.
FEATURES
Location/Qualifiers
source 1..579
/organism="Leishmania major"
/mol_type="genomic DNA"
/strain="Friedlin strain VI"
/db_xref="taxon:5664"
/clone="LMAUFV1_lm71d05"
/lab_host="TOP10 (Invitrogen)"
/clone_lib="Leishmania major FV1 random genomic library"
/note="Vector: pZero-2 (Invitrogen); Site 1: EcoRV;
Genomic DNA was isolated from stationary phase cells. For
this library, DNA was sheared to give a tight size
distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
polymerase, dephosphorylated with Shrimp Alkaline
Phosphatase and ligated into pZero-2 vector's EcoRV site."
BASE COUNT 131 a 135 c 175 g 118 t
ORIGIN
Query Match 2.4%; Score 57.4; DB 28; Length 579;
Best Local Similarity 50.1%; Pred. No. 5.6e-05;
Matches 173; Conservative 0; Mismatches 166; Indels 6; Gaps 1;
QY 1492 CGTAAGCGTGTGCGTGAATGAAGAATATCTCCGATAGCCAGAAATTCGACTG 1551
Db 19 CGAAAGCATCGCGGGCGATGATGAGATGTGCCATGAATCGCTCGAGTACAAAGTC 78
QY 1552 TTGATCAAAAGGCGAGAGGCGTTGAGGTTCTTGCAACCTAGTACGCTACATGGGT 1611
Db 79 CTGAACGGCGCCAGCTTGGCGTGAAGTTAGGCGCAACTCTGTATGCTTCACCGAT 138
QY 1612 TGGAGCGCGCTAGTGTGTTTTCAGGAGTGGCGAAGGCTGTACGGCTGGGGTAGG 1671
Db 139 GCTCAAGGTGCAAGCTACCGTGCCTGAGATCAGTGCCTCCGTCACGGCGATGCGCG 198
QY 1672 CACCTCATAGCAGCCGCTCATCATAGCTGTGTAATAGGCTCAAG-----GTGATC 1725
Db 199 CAGATGATTGACAGGACGAAGAACCTGTTGAGGATCTGTACCCCGCGCGCGTCTG 258
QY 1726 TACGCTGACAGATTCGCTCTTGTGACCTATGATCCGAGAGGTTGGAATTTTCATC 1785
Db 259 TACGCGATCTGACTCTGTGATGTTGTTGTCACCGAGAGGCGGAGCAAG 318
QY 1786 AAAATTATTAAGGAGGAGCTGGGTTTCGAAATCAAGCTAGAGAAG 1830
Db 319 GAGCGCTGCAGGAAGCATGATTTTGGCATCGAGCGCGGAG 363
CNS06CTM 1053 bp DNA linear GSS 14-JUN-2001
T7 end of clone AROAA08A01 of library AROAA from strain CBS 732 of
Zygosaccharomyces rouxii, genomic survey sequence.
RESULT 3
CNS06CTM
LOCUS
DEFINITION
Zygosaccharomyces rouxii, genomic survey sequence.
ACCESSION
AL392960
VERSION
AL392960.1 GI:12142792
KEYWORDS
GSS.
SOURCE
Zygosaccharomyces rouxii
Zygosaccharomyces rouxii
Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Zygosaccharomycetes;
Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes.
REFERENCE
1 (bases 1 to 1053)
Souci, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
Solotkin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durand, P., Lepingle, A., Llorente, B.,
Maupertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissbach, J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of

```

yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)
 20584711
 11152876
 2 (bases 1 to 1053)
 de Montigny, J., Straub, M., Potier, S., Tekai, F., Dujon, B.,
 Wincker, P., Ariguenave, F. and Souciet, J.
 Genomic exploration of the hemiascomycetous yeasts: 8.
 Zygosaccharomyces rouxii
 FEBS Lett. 487 (1), 52-55 (2000)
 20584718
 11152883
 3 (bases 1 to 1053)
 Genoscope.
 Direct Submission
 Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 this GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazzii, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.
 keywords
 Location/Qualifiers
 1..1053
 /organism="Zygosaccharomyces rouxii"
 /mol_type="genomic DNA"
 /strain="CBS 732"
 /db_xref="taxon:4956"
 /clone="AR0AA008A01"
 /clone_lib="AR0AA"
 /note="end : 17"
 <26..>1051
 /note="similar to Saccharomyces cerevisiae ORF YNL102w [
 POL1 ; DNA-directed DNA polymerase alpha, 180 kD subunit
]"
 /evidence=not_experimental
 <398..>1042
 /note="similar to Saccharomyces cerevisiae ORF YDL102w [
 CDC2 ; DNA-directed DNA polymerase delta, catalytic 125 kD
 subunit]"
 /evidence=not_experimental
 BASE COUNT 339 a 182 c 219 g 311 t 2 others
 ORIGIN
 Query Match 2.4%; Score 57.4; DB 29; Length 1053;
 Best Local Similarity 47.3%; Pred. No. 8.5e-05;
 Matches 169; Conservative 1; Mismatches 187; Indels 0; Gaps 0;
 QY 1483 TTGAGCTTCGTAAGCGTGGCTGCTGAAATGAAGAAGTATCTCCGGATAGCCAGAA 1542
 DB 623 TTATGCAACGTCGTCGTAAGTGAAGAAGTATCTGCTACTGAAAGGATCTCTAAG 682
 QY 1543 TATGACTGTTGGATGAAGCGGAGGAGGTTGAAAGTTCTTTCGAAACGCTAGTACGGC 1602
 DB 683 CGTACCAATGTGACATTAAACAGCAGCGGTTAAAGCTAACCGCACTCCATGACGGT 742
 QY 1603 TACATGGGTTGAGCGGCGCTAGTGGTATTCGAGGAGTGCGGAAGGCTGTACCGCT 1662
 DB 743 TGTGTTGGTTATGTAATAGTAGATTTTATGCAACCTTTAGCCATCTTGTGACCAAC 802
 QY 1663 TGGGGTAGGCATCATAGCACCCTCATCATAGCTGCTTAAGTACCTAGGCTCAAGGTG 1722
 DB 803 AAAGTCGTGAGATTTTATGTAATCAAGGCAATTTGGCTGAAAGTTTGAAGACTTACTGTG 862
 QY 1723 ATCTACGGTGACACAGATTCGCTCTTCGTGACCTATGATCCGAGAGGTCGAAATTC 1782
 DB 863 GTATACGGTGATACAGATTCGGTCTATGATGACAGTAATGTGTAATATATCCGATGCT 922

1783 ATCAAAATTATAAAGGAGGAGCTGGGTTCGAATCAAGCTAGAGAGGTGTACAAA 1839
 DB 923 ATTAAGTAGGAAACGAGTTTAAAGAGTTGCTCATGCGCTACAAATKTTAGAA 979

RESULT 4
 C16083/c
 LOCUS
 DEFINITION
 C16083 Clontech human aorta polyA+ mRNA linear EST 30-SEP-1996
 Clone GEN-210D07 5', mRNA sequence.
 C16083
 VERSION
 C16083.1 GI:1570790
 KEYWORDS
 EST.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 483)
 Fujiiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M.,
 Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Takaichi,
 A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H.,
 Shin, S. and Nakamura, Y.
 Fujiiwara et al. (1995)
 Unpublished
 Contact: Tsutomu Fujiiwara
 Otsuka GEN Research Institute
 Otsuka Pharmaceutical Co., Ltd
 463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
 Tel: 0886-65-2888
 Fax: 0886-37-1035.
 Location/Qualifiers
 1..483
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GEN-210D07"
 /clone_lib="Clontech human aorta polyA+ mRNA (#6572)"
 BASE COUNT 120 a 104 c 78 g 177 t 4 others
 ORIGIN
 Query Match 2.2%; Score 54; DB 14; Length 483;
 Best Local Similarity 47.8%; Pred. No. 0.00047;
 Matches 153; Conservative 1; Mismatches 166; Indels 0; Gaps 0;
 QY 85 GAGACGGCGGAGAGGGTAGTCTCTCATGACAGTCTTTTCGCCCATCTTCTATGGCTG 144
 DB 373 GAGACGGGATTTAAGATAGAGCATGATGACATTTTAGACCATACATTTACGCTCT 314
 QY 145 CTTCACCGGCGCGGATCTCTAAGCAGGTAGCACACAGTATTGTCGTCATGATAGGCCA 204
 DB 313 CTCAGGATGATTTCAAGATTGAAGATTGAAGATTGAAGATTGAAGATTGAAGATTGA 254
 QY 205 AAGAGCCGATTTAGGTGTAGGATGACAGAGGAGTACTTTCGGAGGCTCTGAGG 264
 DB 253 ATTGTGAGATTTGATGTAGAGAGGTTGAGAAAGTTTCTCGCAGGCTATTACC 194
 QY 265 GTCTTACGATTCGCACCGCTGCTACCGAGGCTGTTAGGAGTATCGCAACTCGTAAAG 324
 DB 193 GTGTGAAACTTTATTTTGAACATCCCAAGATGTTCCCACTATTAGAGAAAGTTAGA 134
 QY 325 AAGGTTGATGGTGTGAGGATGTTTGTAGAGGCGGATATACGCTTCGCTATGCGCTATC 384
 DB 133 GACATCCAGCACTTTGGGACATCTTCGAATACATATTCATTTCCATTTGCAAGAGATAC 74
 QY 385 ATAGATCACGATCTATTTC 404
 DB 73 ATCGACAAAGCGCTAATACC 54

AL892989 659 bp mRNA linear EST 16-SEP-2002
 LOCUS

AV988988
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
CONTACT: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satohe@acidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1. 576
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="citb33pl9"
/tissue_type="whole animal"
/dev_stage="tailbud embryo"
/clone_lib="Nori Satoh unpublished cdna library, tailbud embryo"

BASE COUNT 185 a 124 c 113 g 154 t
ORIGIN
Query Match 2.2%; Score 52.4; DB 9; Length 576;
Best Local Similarity 49.0%; Pred. No. 0.0015;
Matches 140; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1541 AATATGACATGTTGGTGAAGGACAGAGCGGTGTAAGGTTCTTGCAAGCGCTAGTACG 1600
DB 263 AATATACACAGTATGATATCCGCAAAAGCATTGAATTAACAGCTACAGCATGTATG 322
QY 1601 GCTACATCGGTTGGAGCGGCGTAGTGTTGCGAGGAGTGGCAAGGCTGTACGG 1660
DB 323 GTTGCTTGGATTTCAGCACTCGAGATTTTACGCCAAACCATAGCTCTTGTAAACA 382
QY 1661 CTTGGGTAGCCTCATAGGACCCCATCAACATAGCTCGTAACTAGGCTCAAGG 1720
DB 383 GGCAGGGTGGTGAATTTCTACTGCACACTAAGAACTTGTGAAAAAATGGGATTTGAAG 442
QY 1721 TGATCTACGGTGACACAGATTTCGCTCTTCGTGACCTATGATCCGAGAAAGTGGAAAT 1780
DB 443 TTATTTATGGTGACACAGACTCAATCATGATTAATCAACAGCAGCTACCTTGGCAATG 502
QY 1781 TCATCAAAATTTAAAGGAGGAGCTGGGGTTCGAAATCAAGCTAGA 1826
DB 503 CATTAACTCGGTAAACCGAGTGAATTCGGAGTGAACAACTATA 548

RESULT 7
BG656636
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AV988988
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
CONTACT: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satohe@acidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1. 576
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="citb33pl9"
/tissue_type="whole animal"
/dev_stage="tailbud embryo"
/clone_lib="Nori Satoh unpublished cdna library, tailbud embryo"

BASE COUNT 185 a 124 c 113 g 154 t
ORIGIN
Query Match 2.2%; Score 52.4; DB 9; Length 576;
Best Local Similarity 49.0%; Pred. No. 0.0015;
Matches 140; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1541 AATATGACATGTTGGTGAAGGACAGAGCGGTGTAAGGTTCTTGCAAGCGCTAGTACG 1600
DB 263 AATATACACAGTATGATATCCGCAAAAGCATTGAATTAACAGCTACAGCATGTATG 322
QY 1601 GCTACATCGGTTGGAGCGGCGTAGTGTTGCGAGGAGTGGCAAGGCTGTACGG 1660
DB 323 GTTGCTTGGATTTCAGCACTCGAGATTTTACGCCAAACCATAGCTCTTGTAAACA 382
QY 1661 CTTGGGTAGCCTCATAGGACCCCATCAACATAGCTCGTAACTAGGCTCAAGG 1720
DB 383 GGCAGGGTGGTGAATTTCTACTGCACACTAAGAACTTGTGAAAAAATGGGATTTGAAG 442
QY 1721 TGATCTACGGTGACACAGATTTCGCTCTTCGTGACCTATGATCCGAGAAAGTGGAAAT 1780
DB 443 TTATTTATGGTGACACAGACTCAATCATGATTAATCAACAGCAGCTACCTTGGCAATG 502
QY 1781 TCATCAAAATTTAAAGGAGGAGCTGGGGTTCGAAATCAAGCTAGA 1826
DB 503 CATTAACTCGGTAAACCGAGTGAATTCGGAGTGAACAACTATA 548

RESULT 7
BG656636
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AL892989 XGC-egg Silurana tropicalis cdna clone TEG9041j14 5', mRNA
sequence.
AL892989
VERSION
AL892989.1 GI:22943540
EST.
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
CONTACT: Taylor R
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TEG9041j14.plkSP6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1. 659
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TEG9041j14"
/dev_stage="egg"
/lab_host="Escherichia coli XLI-blue"
/clone_lib="XGC-egg"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from sug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"
2 others

BASE COUNT 214 a 128 c 151 g 164 t
ORIGIN
Query Match 2.2%; Score 54; DB 9; Length 659;
Best Local Similarity 46.3%; Pred. No. 0.00059;
Matches 177; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 1453 GCCTCTTCAACACAGTCTTGAGAGCGGTGTTAGAGCTTCGTAAAGCGTGTGCTGCTGAA 1512
DB 252 GGGATCTTACCACGAGGATACGGAACTGGTGGAGAGGCGGTCACTGAGCAGTTA 311
QY 1513 ATGAAGAAGTATCTCCGGATAGCCAGAAATATCGACTGTGGATGAAGGAGGAGGCG 1572
DB 312 ATGAAGACGCTGACCTAAACCTGATTTATCTTACAGTATGACATCAGACAAAAGCT 371
QY 1573 TTGAAGGTTCTTGCAACCGCTAGTTACGGCTACATGGTTGAGCGGCGCTAGGTGTTAT 1632
DB 372 TTAAGCTCAGCGCAACAGCATGTATGGTGCCTGGGCTTCTCTACAGCAGGTTCTAT 431
QY 1633 TGCAGGAGGTGCCAAGGCTTCACGGCTTGGGGTAGCCTCATATAGCACCAGCCATC 1692
DB 432 GCCAAACCTTGGCTCGATTGGTTACTCCAGGAGGAGAGATTTTGTGATACCAAG 491
QY 1693 AACATAGCTCGTAACTAGGCTCAAGGTCATCTACGGTGACACAGATTCGCTCTTCGGT 1752
DB 492 GAAATGGTACAGAGATGATCTGAGGTCATATATGAGACACGGATTTCTATCATGATA 551
QY 1753 ACTATGATCCGGAGAGGTGGAAATTTTCATCAAAATTAAGAGAGAGCTGGGGTTC 1812
DB 552 AACACCAATTTGTAATTTAGAGAGGCTTCAAGCTTGGAAACCGGCTGAAAGTGAG 611
QY 1813 GAAATCAGCTAGAGGTTGT 1834
DB 612 GTAARCAATCATACAACTTT 633

RESULT 6

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Xenopodinae; Xenopus.
1 (bases 1 to 595)
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R., Waterston, R. and Wilson, R.
WashU Xenopus EST project, 1999
Unpublished
Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library constructed by N. Garrett, K. Ryan and A.M. Zorn, (Wellcome/CRC Institute). DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 504.

FEATURES

Location/Qualifiers
1..595
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:358072"
/tissue_type="head, stage 30"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Wellcome CRC PRN3 head"
/note="Vector: pBSRN3; Site 1: NotI; Site 2: EcoRI; cDNAs were oligo-dT primed and directionally cloned. Staging according to Nieukoop and Faber. Library was constructed by N. Garrett, K. Ryan and A.M. Zorn, (Wellcome/CRC Institute)."

BASE COUNT 193 a 119 c 130 g 153 t
ORIGIN

Query Match 2.2%; Score 52.4; DB 10; Length 595;
Best Local Similarity 46.1%; Pred. No. 0.0016;
Matches 176; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

QY 1453 GGCTTCTCAAGACAGTCTTGAGAGGCTTTAGAGCTTCGTAAAGCGTGGCGTCTGAA 1512
DB 161 GGGATCTTCGAGAGATCGGAACTGGTGAGAGCGCGCTCAGTGAACACTTT 220
QY 1513 ATGAGAGATATCTCCGGATAGCCGAGAAATATGACTGTGGATGAAAGCAGAGCGG 1572
DB 221 ATGAGACGCTGACCTAAACCTGATTTATATCTTGCAGATATGACATCAGACAAAAGCT 280
QY 1573 TTGAGAGTCTTCGAAACCTAGTTACGCTACATGGGTGGAGCGGCTAGGTGTAT 1632
DB 281 CTGAAGCTCAGAGCAATAGCATGTACGATGCTTGGGCTTTTCTACAGCGGTTTAT 340
QY 1633 TGCAGGAGTGGCGAAAGCTGTACGGCTTGGGTAGGACCTCATACGACCGCCATC 1692
DB 341 GCCAACCCTTGGCTGGTGGTGTACTACCAAGGAGAGAGATTTGCTGCATACCAAG 400
QY 1693 ACATAGCTCGTAAACTAGGCTCAAGGTGATCAGGTGACACAGATTCCTTCTGGTG 1752
DB 401 GAAATGGTCAAAAGATGAATCTGGAAGTCATATATGAGACACCGATTCATCATGATA 460
QY 1753 ACCTATGATCCGAGAGAGTGGAAATTTTCATCAAAATTTATAAGAGGAGAGCTGGGTTT 1812
DB 461 AACACTAATGTATATTTAGAGAGTCTTCAGAGCTCGGAAACCGGTAAGAGTGAAG 520
QY 1813 GAATCAAGCTAGAGAGTGT 1834
DB 521 ATAAACAATCATACAAGCTGT 542

RESULT 8

C16480/c

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

C16480
C16480 Clontech human aorta polyA+ mRNA linear EST 30-SEP-1998
clone GEN:330803 5', mRNA sequence.
C16480
C16480.1 GI:1571187
EST.
Homo sapiens (human)
Homo sapiens
Mammalia; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi; Eumalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 389)
Fujiwara, F., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shimomiyu, H., Takachi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H., Shin, S. and Nakamura, Y.
Fujiwara et al. (1995)
Unpublished
Contact: Tsutomu Fujiwara
Otsuka GSN Research Institute
Otsuka Pharmaceutical Co., Ltd
463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035.

FEATURES

Location/Qualifiers
1..389
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GEN:330803"
/clone_lib="Clontech human aorta polyA+ mRNA (#6572)"

BASE COUNT 93 a 91 c 63 g 140 t
ORIGIN

Query Match 2.2%; Score 52; DB 14; Length 389;
Best Local Similarity 48.3%; Pred. No. 0.0015;
Matches 142; Conservative 1; Mismatches 151; Indels 0; Gaps 0;

QY 111 TGACAGGCTTTTCGCCCATCTTCTATGCTGCTGTCACCGCGCGCCATCTTAGCA 170
DB 347 TGRTAGAACTTTTACCATCTATTCAGCHCTCTCAGGATGATTCAAGATTGAAGA 288
QY 171 GGTAGCACACAGTATTCGTGATGAGTAGGAGCCGAGAGCCGATTTATAGTGTAGAGGA 230
DB 287 AGTTAAGAAATAACGGGGGAAAGCATGCAAGATTGTGAGATTGTTGATGTAGAGAA 228
QY 231 TGACAGAGGAGTACTTCGGAGGCTCTGAGGCTCTTACGTTATTCGACCGTGTACC 290
DB 227 GGTGAGAAAGATTTTCGGCAAGCTTATACCGTGTGGAACCTTTTATTTGGAACATCC 168
QY 291 CGAGGCTGTAGGGAGTATCGGAACCTGTAAGAGAGCTTGTGATGTTGAGGATGTTCT 350
DB 167 CCAAGATGTTCCACTATTTAGAGAAAGTTAGAGACATCCAGAGTGTGGACATCTT 108
QY 351 AGAGCGGATATACGCTTCGCTATGCGCTATCTCATAGATCAGCATTTATTC 404
DB 107 CGAATACATATTTCCATTTCGAAAGAGATGATCTCATGACAAAGGCTTAATACC 54

RESULT 9

A1211172/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE

A1211172
A1211172 Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library Emericella nidulans clone oob09a1 3', mRNA sequence.
A1211172
A1211172.1 GI:3773114
EST.
Emericella nidulans (anamorph: Aspergillus nidulans)
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; Emericella.
1 (bases 1 to 352)

AUTHORS Kupfer,D., Gray,J., Hausner,J., Lai,H., Martin,W., Aramayo,R.,
Prade,R. and Roe,B.
TITLE An Aspergillus nidulans EST Database
JOURNAL Unpublished
COMMENT Other ESTs: oob09a1.r1
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: M13-20.
FEATURES Location/Qualifiers
1..352
/organism="Emmericella nidulans"
/mol_type="mRNA"
/strain="FGSC A26"
/db_xref="taxon:162425"
/clone="oob09a1"
/tissue_type="vegetative mycelia, asexual structures"
/clone_lib="Aspergillus nidulans 24hr asexual
developmental and vegetative cDNA lambda zap library"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
3' end of cDNA cloned into XhoI site of pBluescript"

BASE COUNT 75 a 91 c 84 g 102 t
ORIGIN
Query Match 2.1%; Score 51.8; DB 9; Length 352;
Best Local Similarity 47.9%; Pred. No. 0.0016;
Matches 149; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 1897 ATAGATATTGTCGGTTTCAGGCTGTACGTGGCGATTGGTGTGAATCCGCCAAGAGGTT 1956
DB 320 ATGATACGAAGGTTATGAGCTGTTCGACAGACAACTGCTGTCTGCGAAAAACGTG 261
QY 1957 CAGACTAAGTTTGCATATAGTATTGACAGCGGTGAGTGAACAGGCTGTAGAGTAC 2016
DB 260 ATTGAACCGTGTGCACAAATCTTGTATGACCGTGAACCGCTCAAGATAT 201
QY 2017 GTCCAGGAAGTTGTGAAGAGTTGGAGGCGGCAAGTTCCCATAGAGAAGCTTGTATC 2076
DB 200 GTCAAGACACCATATCCGACCTTTTCAGAACACAGATCGACATGTCGAGCTGTGTATC 141
QY 2077 TGGAGACCTTATAGTACGCTCTTGAGGAGTACACAAACGAGGACACACACGTCGTTGCA 2136
DB 140 ACGAAGCTCTGACCAAGGATTCCTATACCGCCAAACAGCTCACCGTTGAGCTCGCAGAA 81
QY 2137 GCGAAGGATGCTGTACAGAGCTACCGGTTAAGCCCGAGCGCAGAGATAGGATGTA 2196
DB 80 CGGATGCGAAGCGTGTATGCGGGTTCGCTCTACGCTAGGCGACCGTGTGCTACGTT 21
QY 2197 ATAGTGAAGG 2207
DB 20 ATTGTCAAAG 10

RESULT 10
AI211173 379 bp mRNA linear EST 19-OCT-1998
LOCUS oob09a1.x1 Aspergillus nidulans 24hr asexual developmental and
DEFINITION vegetative cDNA lambda zap library Emmericella nidulans cDNA clone
oob09a1 5', mRNA sequence.
AI211173
ACCESSION AI211173
VERSION AI211173.1 GI:3773115
KEYWORDS EST.
SOURCE Emmericella nidulans (anamorph: Aspergillus nidulans)
ORGANISM Emmericella nidulans
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Emmericella.

REFERENCE 1 (bases 1 to 379)
AUTHORS Kupfer,D., Gray,J., Hausner,J., Lai,H., Martin,W., Aramayo,R.,
Prade,R. and Roe,B.
TITLE An Aspergillus nidulans EST Database
JOURNAL Unpublished
COMMENT Other ESTs: oob09a1.f1
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: T3.
FEATURES Location/Qualifiers
1..379
/organism="Emmericella nidulans"
/mol_type="mRNA"
/strain="FGSC A26"
/db_xref="taxon:162425"
/clone="oob09a1"
/tissue_type="vegetative mycelia, asexual structures"
/clone_lib="Aspergillus nidulans 24hr asexual
developmental and vegetative cDNA lambda zap library"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
3' end of cDNA cloned into XhoI site of pBluescript"

BASE COUNT 114 a 89 c 97 g 79 t
ORIGIN
Query Match 2.1%; Score 51.8; DB 9; Length 379;
Best Local Similarity 47.9%; Pred. No. 0.0017;
Matches 149; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 1897 ATAGATATTGTCGGTTTCAGGCTGTACGTGGCGATTGGTGTGAATCCGCCAAGAGGTT 1956
DB 52 ATGATACGAAGGTTATGAGCTGTTCGACAGACAACTGCTGTCTGCGAAAAACGTG 111
QY 1957 CAGACTAAGTTTGCATATAGTATTGACAGCGGTGAGTGAACAGGCTGTAGAGTAC 2016
DB 112 ATTGAACCGTGTGCACAAATCTTGTATGACCGTGAACCGCTCAAGATAT 171
QY 2017 GTCCAGGAAGTTGTGAAGAGTTGGAGGCGGCAAGTTCCCATAGAGAAGCTTGTATC 2076
DB 172 GTCAAGACACCATATCCGACCTTTTCAGAACACAGATCGACATGTCGAGCTGTGTATC 231
QY 2077 TGGAGACCTTATAGTACGCTCTTGAGGAGTACACAAACGAGGACACACACGTCGTTGCA 2136
DB 232 ACGAAGCTCTGACCAAGGATTCCTATACCGCCAAACAGCTCACCGTTGAGCTCGCAGAA 291
QY 2137 GCGAAGGATGCTGTACAGAGCTACCGGTTAAGCCCGAGCGCAGAGATAGGATGTA 2196
DB 292 CGGATGCGAAGCGTGTATGCGGGTTCGCTCTACGCTAGGCGACCGTGTGCTACGTT 351
QY 2197 ATAGTGAAGG 2207
DB 352 ATTGTCAAAG 362

RESULT 11
AZ047964 425 bp DNA linear GSS 25-MAY-2001
LOCUS LMAJFV1 lm71d04.x1 Leishmania major FV1 random genomic library
DEFINITION Leishmania major genomic clone LMAJFV1 lm71d04 3', similar to
SW:DFOD SOYBN 04801 DNA POLYMERASE DELTA CATALYTIC CHAIN ;,
genomic survey sequence.
AZ047964
ACCESSION AZ047964
VERSION AZ047964.1 GI:7257910
KEYWORDS GSS.
SOURCE Leishmania major
ORGANISM Leishmania major

Db	336	GTGATGGTGAAGTGTGTCACCGACGAGAGGGCGACGCAAGGAGCGCCTCTCAGGAAGCC	395
Qy	1804	CTGGGGTTCCGAATCAAGCTAGAGAAG	1830
Db	396	ATGGAATTTGGCATCGAGGGCGGCGGAG	422
RESULT 12			
CD283171			
LOCUS	G38615.74	NCI_CGAP_ZK1d1	Danio rerio cDNA clone IMAGE:6523915 5', linear mRNA 674 bp
DEFINITION			
ACCESSION	CD283171		
VERSION	CD283171.1		GI:31060947
KEYWORDS	EST.		
SOURCE	Danio rerio		(zebrafish)
ORGANISM	Danio rerio		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.		
REFERENCE	1	(bases 1 to 674)	
AUTHORS	Amundsen, C., Cachuela, N., Chen, F., Cheung, L. M., Chong, A., Murray, L., Oliva, J., Park, C., Reyes, J., Yungen, J. and Swimmer, C.		
TITLE	Expressed sequence tags from NCI_CGAP_ZK1d1, a <i>Danio rerio</i> kidney library		
JOURNAL	Unpublished		
COMMENT	Contact: Chen F. Exelbros, Inc. 170 Harbor Way, PO Box 511, South San Francisco, CA 94083-0511, USA		

```

Tel: 650 837 7000
Fax: 650 837 8300
Email: fchen@exelixis.com
DNA Sequencing by: Exelixis, Inc. Clone distribution information
can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: 14115 row: C column: 19
High quality sequence stop: 674.
Location/Qualifiers
1. 674
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:6523915"
/lab host="DH10B (TI-resistant)"
/clone lib="NCI_CGAP_EKidi"
NOTE=Organ: Kidney; Vector: pCMV-SPORT6.cdbb; Site 1:
EcorV; Site 2: NotI; Cloned unidirectionally. Primer:
Oligo RT. Average insert size 1.8 kb. Constructed by J.
Wang (Research Genetics, Invitrogen Corp) from tissue
donated by L. Zon (Harvard University). Note: this is a
NCI_CGAP library."
BASE COUNT 213 a 146 c 165 g 150 t
ORIGIN
Query Match 2.1%; Score 50.4; DB 14; Length 674;
Best Local Similarity 48.6%; Pred. No. 0.0065;
Matches 138; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
Qy 1549 CTGTGGATGAAGGCAGACGCGGTTCAGGTTCTTGCAACGCTAGTTCGGCTACATG 1608
Db 15 CAGTATGACATCAGGCAAAAGGCCCTGAAGCTGACAGCTAACAGCATGTAGCGATGCGCTT 74
Qy 1609 GGTGTGGACGGCGGTGTAGTGGTATGTGACGGAGTGGCCAAAGGCTGTCAAGCTTGGGGT 1668
Db 75 GACTTTTCTTCAGTGGCTTTTACGCCAAGCCCTTGGCAGCATTTGGTCAACATAAAGGA 134
Qy 1669 AGGCACCTCATCGCACGGCATCAACATAGCTCGTAAACTAGGCTCAAGGTGATCTAC 1728
Db 135 AGAGAGATTTTGATGTGCACCAAGAAGAAATGGTACAGAGGATGAATCTCGAGGTCAATAC 194
Qy 1729 GGTGTACACAGATTCGTCTCTTGTGTACCTCATGTATGATCCGGAGAGGTTGAAAAATTTTCATCAA 1788

```


Db 195 GGGGACACTGCTCCATCATCATGATTAAACACACAGACCAATTTAGAGAGGCTCTTCAA 254

QY 1789 ATTATTAAGGAGAGGCTGGGGTTCGAAATCAAGCTAGAGAAGGT 1832

Db 255 CTGGGAACAGAGGTCAAAAGTGAGGTGATTAACCTCTATAGCT 298

RESULT 13

CA854286

LOCUS

DEFINITION

710 bp mRNA linear EST 16-DEC-2002
 STRO0656 oligo dt primed shield stage Danio rerio cDNA clone CB578
 5' similar to DNA polymerase alpha catalytic subunit (BC 2.7.7.7),
 mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Danio rerio (zebrafish)
 Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
 1 (bases 1 to 710)

REFERENCE

AUTHORS

Thiisse B., Pflumio, S., Heyer, V., Degraeve, A., Woehl, R., Lux, A.,
 Alunni, V., Agathon, A., Furthauer, M. and Thiisse, C.
 Expression of the zebrafish genome during embryogenesis (11.5.02)

JOURNAL

COMMENT

Unpublished
 Contact: Thiisse B
 Institut de Genetique et de Biologie Moleculaire et Cellulaire
 CNRS, INSERM, ULP
 1, rue Laurent Fries, BP163, CU de Strasbourg, 67404 Illkirch Cedex
 1, France

Tel: 33 3 88 65 33 60

Fax: 33 3 88 65 32 01

Email: thiisse@bmc.u-strasbg.fr

EST from a cDNA of a gene whose expression is spatially restricted
 during embryogenesis. We have established its expression pattern
 during embryonic development by whole mount in situ hybridization
 on zebrafish embryos from the gastrula stage to 2 days of
 development. The corresponding data are available on the zebrafish
 community database at <http://zfinfo.org/> cDNA library preparation:
 Matthew Clark (Lehrach lab; Max Planck Institut fuer Molekulare
 Genetik, Berlin). (Vector: pSPORT1; Site 1: NotI; Site 2: SalI; 1st
 strand cDNA was primed with a Not I - oligo (dT)15 primer
 [5'-GCACGTCTGATGCGGAGCGGCGCCCTTTTCTTTTCTTTT3'];

double-stranded DNA was ligated to Sal I adaptors (BRL), digested
 with Not I and cloned into the Not I and Sal I sites of the pSPORT1
 vector (BRL). DNA sequencing by: ICBMC sequencing facility. Clone
 distribution: zebrafish international resource center at the
 University of Oregon (Institute of Neuroscience, 1254 University of
 Oregon, Eugene, OR 97403-1254)
 Seq primer: 17 TAATACGACTCACTATAGG.

FEATURES

source

1..710

Location/Qualifiers

/organism="Danio rerio"

/mol_type="mRNA"

/db_xref="taxon:7955"

/clone="CB678"

/dev_stage="shield stage"

/clone_lib="oligo dt primed shield stage"

/note="Vector: pSPORT1 (BRL); Site 1: NotI; Site 2: SalI;

Oligo dt cDNA library constructed from RNA of shield

stage zebrafish embryos"

232 a 150 c 179 g 149 t

BASE COUNT

ORIGIN

Query Match 2.1%; Score 50.4; DB 14; Length 710;
 Best Local Similarity 48.6%; Pred. No. 0.0068;
 Matches 138; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1549 CTGTTGGATGAAGGAGGCGTTGAAAGGTTCTTTCGAAACGCTAGTTACGGCTACATG 1608

Db 172 CAGTATGACATCAGGCAAAAGGCCCTGAAAGCTGACAGCTAACAGCATGTACGGATGCCTT 231

QY 1609 GGTTCGAGCGCGCTAGTGTGTATTGTCAGGAGTGCACAAAGCGTGTCAAGCGCTTGGGGT 1668

Db 232 GGCCTTTCTTTCAGTCGCTTTTACGCCAAGCCTTTGGCAGCATTTGGTACGCATAAGGA 291

QY 1669 AGGCACCTCATACGACCGCCATCAATAGCTCGTAAACTAGGCTCTCAGGGTATCTAC 1728

Db 292 AGAGAGATTTTGTATGTCACACCAAGAAATGTCACAGAGATGATCTCTGGAGGTCATATAC 351

QY 1729 GGTGACACAGATTCGCTCTTCTGTCACCTATGATCCGGAGAGGTTGGAAATTTTCATCAA 1788

Db 352 GGGACACTGACTCATCATCATGATTAACACACAGCACCATTATAGAGGAGGCTTCTCAA 411

QY 1789 ATTATAAGGAGGAGCTGGGTTCTCGAAATCAAGCTAGAGAGGT 1832

Db 412 CTGGAAACAAGGTCAAAAGTGAGGTGAATAAACTCTATAAGCT 455

RESULT 14

BI889706

LOCUS

DEFINITION

752 bp mRNA linear EST 12-OCT-2001
 ZF637-2-001461 Zebrafish shield stage whole embryo cDNA library
 MPMPG637 Danio rerio cDNA clone MPMPG637_3J3, MPMPG637J033 5', mRNA
 sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Danio rerio (zebrafish)
 Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
 1 (bases 1 to 752)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Hennig S
 Laboratory 123, Dept. Lehrach
 Max-Planck-Institut fuer Molekulare Genetik
 Ihnestr. 63-73, D-14195 Berlin, Germany
 Tel: +49 30 8413 1612
 Fax: +49 30 8413 1380
 Email: hennig@molgen.mpg.de
 5' EST sequencing of clones from a zebrafish shield stage library,
 normalised from 55,000 starting clones by oligonucleotide
 fingerprinting
 High quality sequence stop: 752.

FEATURES

source

1..752

Location/Qualifiers

/organism="Danio rerio"

/mol_type="mRNA"

/db_xref="taxon:7955"

/clone="MPMPG637_3J3;MPMPG637J033"

/tissue_type="whole embryo"

/dev_stage="shield stage, 6 hrs post-fertilisation"

/lab_host="E.coli, Xli blue MRF"

/clone_lib="Zebrafish shield stage whole embryo cDNA
 library MPMPG637"

/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI;

oligo-dT-NotI primed, SalI adaptors, directionally cloned,
 library normalised by oligonucleotide fingerprinting"

240 a 172 c 189 g 151 t

BASE COUNT

ORIGIN

Query Match 2.1%; Score 50.4; DB 12; Length 752;
 Best Local Similarity 48.6%; Pred. No. 0.007;
 Matches 138; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1549 CTGTTGGATGAAGGAGGCGGTTCGAAAGTTCCTTTCGAAACGCTAGTTACGGCTACATG 1608

Db 183 CAGTATGACATCAGGCAAAAGGCCCTGAAAGCTGACAGCTAACAGCATGTACGATGCCTT 242

QY 1609 GGTTCGAGCGCGCTAGTGTGTATTGTCAGGAGTGCACAAAGCGTGTACGGCTTGGGGT 1668

1585 GCAAAACGCTAGCTTACGGCTACATCGGTTGGAGCGCGCTAGTGGTATTTGAGGGAGTGC 1644
216 GCAAAACGCTATGATGTTGGCTGGGATTTTGGTACAGCAGATTTTATGCTAAACCACTG 275
1645 GCAAAAGGCTGTACCGGCTTGGGTAGGACCTCATACGCCCGCCATCAACATAGCTCGT 1704
276 GCTGCCCTTGGTGACATACAAAGGAAGGAGATTTTGTATGCTATACGAAAGACATGGTACAG 335
1705 AAACCTAGGCTCAAGGTCATCTACGGTGACACAGATTCGCTCTTCTGTCGACCTATGATCCG 1764
336 AGATGAATCTTGAAGTTATTTTATGAGACACAGATTCATATGATAAATACCAATAGC 395
1765 GGAAGAGGTGGAAAATTTTCATCAAAATTTATTAAGGAGG 1801
396 ACCAACCTGGAGAAGTATTTAAGTTGGGAAACAGG 432

Search completed: November 26, 2003, 13:48:02
Job time : 4966 secs

243 GGCTTTTCTTCAGTCGCTTTTACGCCAAGCCCTTGGCAGCATTTGGTACGCTAAAGGA 302
1669 AGGCACTCATACGACCGCCATCAACATAGCTCGTAAACTAGGCTCAAGGTGATCTAC 1728
303 AGAGAGATTTTGTATGACACACCAAGAAATGGTACAGAGGATGAATCTGGAGGTCTATAC 362
1729 GGTGACACAGATTCGCTTCTGTCGACCTATGATCCGAGAGGTGGAATTTTCATCAA 1788
363 GGGGACACTGACTCCATCATGATTAACACCAACAGCACCATTATGAGGAGGTCTTCAA 422
1789 ATTATAAGCAGAGGCTGGGGTTCGAAATCAAGCTAGAGAGGT 1832
423 CTGGGAAACAAAGGTCAAAAGTGAGGTGATTAATTAAGTCTTAAAGT 466

RESULT 15
BF136910
LOCUS 601782303F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4010112 5',
DEFINITION mRNA sequence.
ACCESSION BF136910
VERSION BF136910.1 GI:10975950
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 922)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L2AM9247 row: f column: 01
High quality sequence stop: 680.
Location/Qualifiers
1..922
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:4010112"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/clone_lib="NCI CGAP Lu30"
/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; transgenic model WNT-1, expression driven by
WNTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
dT. Library constructed by Life Technologies.
Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 299 a 172 c 241 g 210 t
ORIGIN

Query Match 2.0%; Score 49; DB 10; Length 922;
Best Local Similarity 46.6%; Pred. No. 0.021;
Matches 157; Conservative 0; Mismatches 180; Indels 0; Gaps 0;
QY 1465 ACAGTTCTTGAGAGGCTGTAGAGCTTCGTAAGCGTGGCTGCTGAATGAAGAAGTAT 1524
Db 96 AGAGAGATTCGGAACCTGGTAGAGCGGAGAAACAGTCAACAGCTATGAAACAGCAA 155
QY 1525 CTCCCGGATAGCCAGATATCGACTGTTGGATGAAGCAGAGCGGTTGAAGTCTTT 1584
Db 156 GATTTAAATCCAGACCTTGTCTTCAGTATGATATCCGACAGAGGCTTTGAAGCTACA 215

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2003, 10:18:29 ; Search time 629 Seconds

(without alignments)

10351.422 Million cell updates/sec

Title: US-10-034-849-1

Perfect score: 2412
Sequence: 1 atgactgaagtgtattccac.....tagccaagaagaagtaa 2412

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 19Jun03.*
1: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDs1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2412	100.0	2412	24	Pyrolobus fumaria
2	2412	100.0	2412	25	DNA encoding Pyrol
3	2412	100.0	2412	25	DNA encoding Pyrob
4	2410.4	99.9	2412	20	P. fumarius DNA po
5	2400.8	99.5	2412	24	Pyrolobus fumaria
6	2400.8	99.5	2412	25	DNA encoding Pyrob
7	2338	96.9	2405	24	Pyrolobus fumaria
8	1071.2	44.4	3068	17	Pyrodictium occult

9	1069.6	44.3	2430	15	AAQ73844
10	1045.6	43.3	2430	15	AAQ73843
11	580	24.0	2610	21	AA50678
12	279	11.6	2367	20	AA22284
13	279	11.6	2367	24	AA33183
14	279	11.6	2367	25	ABX14887
15	204.2	8.5	2376	18	AA288374
16	200.8	8.3	3731	18	AA288373
17	193.2	8.0	422	17	AA216278
18	191.4	7.9	2325	19	AA214524
19	189.8	7.9	2325	19	AA223352
20	188	7.8	2325	18	AA286434
21	184.8	7.7	2325	22	AA170053
22	177.4	7.4	2322	24	ABL56547
23	174.4	7.2	2374	19	AAV05805
24	174.4	7.2	2374	19	AAV05807
25	167.2	6.9	265118	22	AAH41227
26	163.2	6.8	2322	19	AAV11396
27	162.6	6.7	4446	19	AAV36034
28	162.2	6.7	2995	19	AAV36035
29	161.6	6.7	2322	19	AAV26371
30	146.2	6.1	2734	17	AA238567
31	132	5.5	2328	21	AA289727
32	131.4	5.4	5340	17	AA210258
33	131.4	5.4	5340	18	AA271296
34	131.4	5.4	5342	17	AA228360
35	131.4	5.4	5342	17	AA214694
36	131.4	5.4	5342	19	AAV14525
37	131.4	5.4	5342	24	AA223351
38	123	5.1	5255	22	AA505831
39	121.8	5.0	3799	18	AA288384
40	110	4.6	2328	19	AAV66490
41	110	4.6	2328	22	AA290292
42	110	4.6	2535	24	AA24821
43	110	4.6	3499	17	AA242154
44	110	4.6	3499	17	AA28282
45	110	4.6	3499	20	AA200392

ALIGNMENTS

RESULT 1

AD35187 standard; DNA; 2412 BP.

AD35187;

25-JUL-2002 (first entry)

Pyrolobus fumaria DNA polymerase encoding DNA, 1PY2.

DNA polymerase; thermostable; enzyme; gene; ds.

Pyrolobus fumaria.

Key Location/Qualifiers

CDS 1..2412

/*tag= a

/product= "Pyrolobus fumaria DNA polymerase, 1PY2"

WO200220735-A2.

14-MAR-2002.

06-SEP-2001; 2001WO-US28007.

06-SEP-2000; 2000US-0656309.

(DIVE-) DIVERSA CORP.

Callen W, Mathur EU, Short JM;

XX

DR WPI: 2002-362247/39.
 DR P-PSDB; AAE22118.
 XX
 PT New thermostable polymerase useful for sequencing DNA, amplifying
 PT double stranded DNA, or incorporating a non-natural nucleotide or a
 PT nucleotide analog into a DNA molecule -
 XX
 PS Claim 1; Fig 1A-E; 161pp; English.
 XX
 CC The invention relates to thermostable DNA polymerases having high
 CC temperature polymerase activity, such as those derived from
 CC Pyrobolus fumaria and nucleic acid molecules encoding such polymerases.
 CC Polymerases are useful for catalysing the formation or repair of a
 CC nucleic acid sequence and for modifying small molecules. They are also
 CC useful for sequencing DNA molecules, for preparing cDNA from mRNA, for
 CC amplifying double stranded DNA molecules and for incorporating non-
 CC natural nucleotides or nucleotide analogues into a DNA molecule. The
 CC present sequence is Pyrobolus fumaria DNA polymerase encoding DNA.
 XX
 SQ Sequence 2412 BP; 631 A; 512 C; 707 G; 562 T; 0 other;

Query Match 100.0%; Score 2412; DB 24; Length 2412;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACTGAGTTGTTATTCACGGTTTATGACTCTAGCTACGAGTTGTTGGTAAAGAGCCT 60
 DB 1 ATGACTGAGTTGTTATTCACGGTTTATGACTCTAGCTACGAGTTGTTGGTAAAGAGCCT 60
 QY 61 CAGGTAATCATATGGGGTATTGCTGAGAACGGCGAGAGGTAGTCTCTCATTAACAGAGTCT 120
 DB 61 CAGGTAATCATATGGGGTATTGCTGAGAACGGCGAGAGGTAGTCTCTCATTAACAGAGTCT 120
 QY 121 TTTCGCCATATCTTCATCGCTGTGTCACCGCGCGCGATCTTAACAGAGTACACAA 180
 DB 121 TTTCGCCATATCTTCATCGCTGTGTCACCGCGCGCGATCTTAACAGAGTACACAA 180
 QY 181 CGTATTCGTCATGAGTAGGCGCAAGACCGCGATATAGGTGAGAGGATGACAGAGG 240
 DB 181 CGTATTCGTCATGAGTAGGCGCAAGACCGCGATATAGGTGAGAGGATGACAGAGG 240
 QY 241 AAGTACTTCGGAGGCTCGTAGGGTCTTACGATTCGACCGTGTCTACCCAGGCTGT 300
 DB 241 AAGTACTTCGGAGGCTCGTAGGGTCTTACGATTCGACCGTGTCTACCCAGGCTGT 300
 QY 301 AGGAGTATCGGAACTCGTAAAGACGTTGATGTTGAGTGTGAGAGTGTCTAGAGCGGAT 360
 DB 301 AGGAGTATCGGAACTCGTAAAGACGTTGATGTTGAGTGTGAGAGTGTCTAGAGCGGAT 360
 QY 361 ATACGCTTCGCTATGCGCTATCTCATAGATCAGATCTATTTCCCTTACCTGTTACCGT 420
 DB 361 ATACGCTTCGCTATGCGCTATCTCATAGATCAGATCTATTTCCCTTACCTGTTACCGT 420
 QY 421 GTAGAGGCTGAGCCCTCGAGAACAGATGGGCTTCCGTTGTCGACAGGTATACCTGGTT 480
 DB 421 GTAGAGGCTGAGCCCTCGAGAACAGATGGGCTTCCGTTGTCGACAGGTATACCTGGTT 480
 QY 481 AAGAGCAGGCGGAGGCACTTTATGTTGAGGCTCTCGCACCAACCAAGCTTCCCGATCTT 540
 DB 481 AAGAGCAGGCGGAGGCACTTTATGTTGAGGCTCTCGCACCAACCAAGCTTCCCGATCTT 540
 QY 541 AGGATACCTCGCTTCGATATTAAGTTTATAGCAAGAGGTCGCGGCTCCAGAGCGC 600
 DB 541 AGGATACCTCGCTTCGATATTAAGTTTATAGCAAGAGGTCGCGGCTCCAGAGCGC 600
 QY 601 GATCCCTGTAATAGTAGTAGCTGTGAAGACTGACGATGGCGATGAGTGTCTATTCTTGA 660
 DB 601 GATCCCTGTAATAGTAGTAGCTGTGAAGACTGACGATGGCGATGAGTGTCTATTCTTGA 660
 QY 661 GAGGCAAGACGATCGAAGAACCGATACCGAGTTTGTAGAGTACGTGAAGAGGTATGAC 720
 DB 661 GAGGCAAGACGATCGAAGAACCGATACCGAGTTTGTAGAGTACGTGAAGAGGTATGAC 720

QY 721 CCGACATATAGTCGGTTATTAACAACAATCATTTGATTTGGCTTATCTTTTGGAGCGC 780
 DB 721 CCGACATATAGTCGGTTATTAACAACAATCATTTGATTTGGCTTATCTTTTGGAGCGC 780
 QY 781 GCGCGCATCTAGGCATAAAGCTTGTGATCTAGAGAGTTGGCGCCGAGCCACACT 840
 DB 781 GCGCGCATCTAGGCATAAAGCTTGTGATCTAGAGAGTTGGCGCCGAGCCACACT 840
 QY 841 AGGTATCATGGCAGCGCTCTCTGTCCTGCGGAGGCTTAAAGTAGATCTGTACGATATGCC 900
 DB 841 AGGTATCATGGCAGCGCTCTCTGTCCTGCGGAGGCTTAAAGTAGATCTGTACGATATGCC 900
 QY 901 GAAGAGATGCGCAGAGATCAAGATTAAGAGGCTCTGAGAGGTCGACAGATATCTAGCGGTG 960
 DB 901 GAAGAGATGCGCAGAGATCAAGATTAAGAGGCTCTGAGAGGTCGACAGATATCTAGCGGTG 960
 QY 961 ATGAAGAAGAGTGAACGCGTTATCATCAATTTGTTGGAGATTCAGACTATTTGGAGCGAC 1020
 DB 961 ATGAAGAAGAGTGAACGCGTTATCATCAATTTGTTGGAGATTCAGACTATTTGGAGCGAC 1020
 QY 1021 CCGAAGAGAGACCACTATTACTGCAATACGCGCGAGCGATGTCGCGCTACTTACGGC 1080
 DB 1021 CCGAAGAGAGACCACTATTACTGCAATACGCGCGAGCGATGTCGCGCTACTTACGGC 1080
 QY 1081 TTAGCCGAGAGATATTGCGGTTTGTCTATCCAGTTGCTGATACGTAACAGGCTCTCCACTA 1140
 DB 1081 TTAGCCGAGAGATATTGCGGTTTGTCTATCCAGTTGCTGATACGTAACAGGCTCTCCACTA 1140
 QY 1141 GACAGTAGTGGTGGATGAGTGTGGCTTTTCGACTTGAATGGTACCTGATACGCGGGCG 1200
 DB 1141 GACAGTAGTGGTGGATGAGTGTGGCTTTTCGACTTGAATGGTACCTGATACGCGGGCG 1200
 QY 1201 TTTAAGATGAAGAGCTTGTGCGGACCGGTTGAGCGCCGAGAGAGCTTACCGTGGC 1260
 DB 1201 TTTAAGATGAAGAGCTTGTGCGGACCGGTTGAGCGCCGAGAGAGCTTACCGTGGC 1260
 QY 1261 GCTATAGTCTTTCGAGCGGTTGAGAGGCTGACAGAGATATAGCCGCTACTCGACTTTAGC 1320
 DB 1261 GCTATAGTCTTTCGAGCGGTTGAGAGGCTGACAGAGATATAGCCGCTACTCGACTTTAGC 1320
 QY 1321 TCGATGTACCCAAACATCATGTATTAAGTACATGTTGGTCTCTGACAGCTTTGAGGCT 1380
 DB 1321 TCGATGTACCCAAACATCATGTATTAAGTACATGTTGGTCTCTGACAGCTTTGAGGCT 1380
 QY 1381 GGTCAAAGTGTGGGAGTGTGTTGCTGGAGGCCCGGAGGTCAAGACAGAGTTCCT 1440
 DB 1381 GGTCAAAGTGTGGGAGTGTGTTGCTGGAGGCCCGGAGGTCAAGACAGAGTTCCT 1440
 QY 1441 AGGTGTCCGCCCGCTTTCTTCAAGACAGTTCCTTGGAGGCTGTTAGAGCTTCGTAAGCGT 1500
 DB 1441 AGGTGTCCGCCCGCTTTCTTCAAGACAGTTCCTTGGAGGCTGTTAGAGCTTCGTAAGCGT 1500
 QY 1501 GTGCTGCTGAAATGAAGAGATATCTCCGATAGCCAGATATCAACTGTTGGATGAA 1560
 DB 1501 GTGCTGCTGAAATGAAGAGATATCTCCGATAGCCAGATATCAACTGTTGGATGAA 1560
 QY 1561 AGGCAAGAGGCTTGAAGGTTCTTGCACAACTAGTTACGCTACATGGTGGTGGAGCGC 1620
 DB 1561 AGGCAAGAGGCTTGAAGGTTCTTGCACAACTAGTTACGCTACATGGTGGTGGAGCGC 1620
 QY 1621 GCTAGGTGTTATTCAGGAGTGTGCGAAGGCTGTACGCGCTGGGGTAGGCACCTCAT 1680
 DB 1621 GCTAGGTGTTATTCAGGAGTGTGCGAAGGCTGTACGCGCTGGGGTAGGCACCTCAT 1680
 QY 1681 CGCACCGCATCAACATAGCTCGTAAACTAGGCTCTAAGGTGATCTACGCTGACACAGAT 1740
 DB 1681 CGCACCGCATCAACATAGCTCGTAAACTAGGCTCTAAGGTGATCTACGCTGACACAGAT 1740
 QY 1741 TCGCTCTTCGTGACCTATGATCCGAGAGGTGGAATTTTCATCAAAATTTATAAGGAG 1800
 DB 1741 TCGCTCTTCGTGACCTATGATCCGAGAGGTGGAATTTTCATCAAAATTTATAAGGAG 1800
 QY 1801 GAGCTGGGTTTCGAAATCAAGCTAGAGAGGTGTACAAACGCTTTATTTCTTTACAGAGCT 1860

```

Db 1801 GAGCTGGGTTGGAATCAAGTAGAGAGGTGTACAAACGCTTATCTTACAGAGGCT 1860
Qy 1861 AAGAAGAGGTACGCTGCGCTTCTCGAGGACGCGATATATATGTCGTTTCGAGGCT 1920
Db 1861 AAGAAGAGGTACGCTGCGCTTCTCGAGGACGCGATATATATGTCGTTTCGAGGCT 1920
Qy 1921 GTACGTCGCTGTTGGTGTGAACCTGCCAAGGAGGTTTCAGACTAAGGTTGTGCAAAATAGTA 1980
Db 1921 GTACGTCGCTGTTGGTGTGAACCTGCCAAGGAGGTTTCAGACTAAGGTTGTGCAAAATAGTA 1980
Qy 1981 TTGAAGCAGGTGAGGTGCAACAGGCTGTAGAGTACGTCTAGGAAGATTTGTGAAGAGTTG 2040
Db 1981 TTGAAGCAGGTGAGGTGCAACAGGCTGTAGAGTACGTCTAGGAAGATTTGTGAAGAGTTG 2040
Qy 2041 GAGAGGCGAAGGTTCCCATAGAGAGCTTTGTAATCTGGAAGACCCCTAGTAGAGCTCTT 2100
Db 2041 GAGAGGCGAAGGTTCCCATAGAGAGCTTTGTAATCTGGAAGACCCCTAGTAGAGCTCTT 2100
Qy 2101 GAGGAGTACACAAACGAGGACCAACACGCTGCTTGCAGCGAAGAGATGCTGTAGCAGGC 2160
Db 2101 GAGGAGTACACAAACGAGGACCAACACGCTGCTTGCAGCGAAGAGATGCTGTAGCAGGC 2160
Qy 2161 TACCGGTTAAGCCAGGCGACACAGATAGGGTATGTATAGTGAAGGGTGTGGCCGATC 2220
Db 2161 TACCGGTTAAGCCAGGCGACACAGATAGGGTATGTATAGTGAAGGGTGTGGCCGATC 2220
Qy 2221 AGTCAAGAGATGCGCATCTTCAATGTCGAAGATCTTCAAGATCTTCAAGATCTTCAAGATCT 2280
Db 2221 AGTCAAGAGATGCGCATCTTCAATGTCGAAGATCTTCAAGATCTTCAAGATCTTCAAGATCT 2280
Qy 2281 TATGTTGACCAACCAATCATCCGCTGCAATGAGATATCTTGGCTACTTTGGCATCAC 2340
Db 2281 TATGTTGACCAACCAATCATCCGCTGCAATGAGATATCTTGGCTACTTTGGCATCAC 2340
Qy 2341 GAGAAGAGCTGAAAGCAAGTGCACCTGGGCGAGAGACTCTCTTCGACTTTCTAGCCCAAG 2400
Db 2341 GAGAAGAGCTGAAAGCAAGTGCACCTGGGCGAGAGACTCTCTTCGACTTTCTAGCCCAAG 2400
Qy 2401 AAGGCAAGTAA 2412
Db 2401 AAGGCAAGTAA 2412

```

RESULT 2

ABX14886
ID ABX14886 standard; DNA; 2412 BP.

XX ABX14886;

XX 08-APR-2003 (first entry)

XX DNA encoding *Pyrobolus fumarius* thermostable DNA polymerase protein.

XX Gene; ds; thermostable; DNA polymerase; DNA repair;

XX polymorphism identification.

XX *Pyrobolus fumarius*.

XX Key Location/Qualifiers

XX CDS 1..2412

XX /tag= a

XX /product= "P. fumarius DNA polymerase"

XX US2002132243-A1.

XX 19-SEP-2002.

XX 06-SEP-2001; 2001US-0948369.

XX 06-AUG-1997; 97US-0907166.

XX 07-SEP-1999; 99US-0391340.

XX 06-SEP-2000; 2000US-0656309.

XX (CALL/) CALLEN W.
PA (MATH/) MATHUR E. J.
PA (SHORT/) SHORT J.

XX Callen W, Mathur EJ, Short J;

XX WPI; 2003-182285/18.

XX P-PSDB; ABG73155.

XX Novel DNA polymerases having increased activity and stability at
PT increased pH and temperature, useful for DNA sequencing, amplification
PT and incorporating non-natural nucleotides or nucleotide analogs -

XX Claim 1; Fig 1; 81pp; English.

XX This invention relates to a novel purified polypeptide derived from
CC *Pyrobolus fumarius* which has thermostable DNA polymerase activity. The
CC protein of the invention is useful for catalysing the formation or
CC repair of a nucleic acid sequence, for comparing a sequence to a
CC reference sequence and for identifying polymorphisms. The nucleic
CC acid sequence of the invention is useful for preparing cDNA from mRNA,
CC and in an amplification process of a double-stranded DNA molecule.
CC The nucleotide sequence is also useful for incorporating non-natural
CC nucleotides or its analogues into a DNA molecule, by contacting a
CC polypeptide encoded by the nucleic acid with a DNA template in a PCR
CC amplification reaction. The nucleotides which may be used for this are
CC inosine, 2-aminopurine or 5-methylcytosine. A nucleic acid probe
CC derived from the nucleic acid sequence of the invention is useful in
CC chromosome walking procedures to identify clones containing genomic
CC sequences located adjacent to a sequence of the gene encoding the DNA
CC polymerases of the invention. Such methods allow the isolation of
CC genes which encode additional proteins from the host organisms. The
CC polymerases of the invention have increased activity and stability at
CC increased pH and temperature, and high processivity. The present
CC sequence represents a DNA sequence encoding an *Pyrobolus fumarius*
CC thermostable DNA polymerase protein of the invention.

XX Sequence 2412 BP; 631 A; 512 C; 707 G; 562 T; 0 other;

Query Match 100.0%; Score 2412; DB 25; Length 2412;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 ATGACTGAAGTTGTTATTCACGGTTTATAGACTCTAGCTACGAGGTTCTTGTAAAGAGCT 60
Db 1 ATGACTGAAGTTGTTATTCACGGTTTATAGACTCTAGCTACGAGGTTTGTAAAGAGCT 60
Qy 61 CAGGTAATCATATGGGGTATTGCTGAGAACGGCGAGAGGGTAGTCTCTATTGACAGGTCT 120
Db 61 CAGGTAATCATATGGGGTATTGCTGAGAACGGCGAGAGGGTAGTCTCTATTGACAGGTCT 120
Qy 121 TTTTCGCCCATACCTTCTATGCGCTGCTTGACCGGGCGCGATCCTTAAGCAGGTAGACAA 180
Db 121 TTTTCGCCCATACCTTCTATGCGCTGCTTGACCGGGCGCGATCCTTAAGCAGGTAGACAA 180
Qy 181 CGTATTTCGTGCTATTGAGTAGGCCAAGAGCCGATATATAGGTAGAGATGACAGAGG 240
Db 181 CGTATTTCGTGCTATTGAGTAGGCCAAGAGCCGATATATAGGTAGAGATGACAGAGG 240
Qy 241 AAGTACTTCGGGAGGCGCTCGTAGGGTCTTACGTAATTCGACCGCTGCTACCCGAGGCTGT 300
Db 241 AAGTACTTCGGGAGGCGCTCGTAGGGTCTTACGTAATTCGACCGCTGCTACCCGAGGCTGT 300
Qy 301 AGGAGATATCGCGAACTCGTAAAGAACGTTGATGGTGTGAGATGTTCTTAGAGGCGGAT 360
Db 301 AGGAGATATCGCGAACTCGTAAAGAACGTTGATGGTGTGAGATGTTCTTAGAGGCGGAT 360
Qy 361 ATACGCTTCGCTATGCGCTATCTCATAGATCAGATCTATTTCTTCCCTTACCTGACCGT 420
Db 361 ATACGCTTCGCTATGCGCTATCTCATAGATCAGATCTATTTCTTCCCTTACCTGACCGT 420
Qy 421 GTAGAGGCTGAGCCCTCGAGAACAAAGATGGGCTTCGTTGTCGACAAAGGTATACCTGGTT 480

```

Db 421 GTAGAGCTGAGCCCTCGGACACAGATGGGCTTCGGTGTGACAAAGTATACCTGGTT 480
QY 481 AAGAGCAGCGCGAGCCACTTTATGTGAGGCTCTCGCACCAACCAAGCTTCCCGATCTT 540
Db 481 AAGAGCAGCGCGAGCCACTTTATGTGAGGCTCTCGCACCAACCAAGCTTCCCGATCTT 540
QY 541 AGGATAGCTCGGCTTCGATATGAAGTTTATAGCAACCAAGGCTCGCGGTCGAGAGCGC 600
Db 541 AGGATAGCTCGGCTTCGATATGAAGTTTATAGCAACCAAGGCTCGCGGTCGAGAGCGC 600
QY 601 GATCCTGTAATAGTAGTAGCTGTGAAGACTGACGATGCGCATGAGGTGCTATTCATTGCA 660
Db 601 GATCCTGTAATAGTAGTAGCTGTGAAGACTGACGATGCGCATGAGGTGCTATTCATTGCA 660
QY 661 GAGGCAAGACGATCGAAMACCGATACCGGAGTTGTAGAGTACGTGAAGAGGTATGAC 720
Db 661 GAGGCAAGACGATCGAAMACCGATACCGGAGTTGTAGAGTACGTGAAGAGGTATGAC 720
QY 721 CCGGACATATAGTTCGGTTATACAAACATCATTTTCGATTGGCTTATCTTTTGGGCGC 780
Db 721 CCGGACATATAGTTCGGTTATACAAACATCATTTTCGATTGGCTTATCTTTTGGGCGC 780
QY 781 GCCCGCATCTAGGCAATAGCTTGTATGACTAGAGAGTTGGCGCGAGGCCACCACT 840
Db 781 GCCCGCATCTAGGCAATAGCTTGTATGACTAGAGAGTTGGCGCGAGGCCACCACT 840
QY 841 AGGTCATATGGGCAAGCTCTCTGCTCCCTGGCAGGCTTAACTGATCTGTACGATATGCC 900
Db 841 AGGTCATATGGGCAAGCTCTCTGCTCCCTGGCAGGCTTAACTGATCTGTACGATATGCC 900
QY 901 GRAGATGTCGAGATCAAGATAAAGATCTCGAGGAGTTCGAGAGTATCTAGGCGTG 960
Db 901 GRAGATGTCGAGATCAAGATAAAGATCTCGAGGAGTTCGAGAGTATCTAGGCGTG 960
QY 961 ATGAAGAAGAGTGAACCGGTTATCATCAATTTGGTGGAGATTCACAGACTATTGGGAGC 1020
Db 961 ATGAAGAAGAGTGAACCGGTTATCATCAATTTGGTGGAGATTCACAGACTATTGGGAGC 1020
QY 1021 CCGAAGAGAGACCACTATTACTGCAATACCGCGCGAGATGTCGCGCTACTTACGCG 1080
Db 1021 CCGAAGAGAGACCACTATTACTGCAATACCGCGCGAGATGTCGCGCTACTTACGCG 1080
QY 1081 TTAGCCGAGAGATATTGCGTTTGTCTATCCAGTTCGTACGATACAGTCTCCCACTA 1140
Db 1081 TTAGCCGAGAGATATTGCGTTTGTCTATCCAGTTCGTACGATACAGTCTCCCACTA 1140
QY 1141 GACCAAGTATGCGATGAGTGTGGCTTTTCGACTTGAATGSPACTGATACGCGCGG 1200
Db 1141 GACCAAGTATGCGATGAGTGTGGCTTTTCGACTTGAATGSPACTGATACGCGCGG 1200
QY 1201 TTTAAGTGAAGAGCTTGGCCGAACCGGTTGAGCGCCGAGAGACTTACGTTGCG 1260
Db 1201 TTTAAGTGAAGAGCTTGGCCGAACCGGTTGAGCGCCGAGAGACTTACGTTGCG 1260
QY 1261 GCTATAGTCTTTCGAGCGGTTGAGAGCGGTGCACAGATATAGCCGCTACTCCGACTTAGC 1320
Db 1261 GCTATAGTCTTTCGAGCGGTTGAGAGCGGTGCACAGATATAGCCGCTACTCCGACTTAGC 1320
QY 1321 TCGATGATCCCAACATCATGATAAAGTACAATGTTGGTCTTGACACGCTTGTGAGGCT 1380
Db 1321 TCGATGATCCCAACATCATGATAAAGTACAATGTTGGTCTTGACACGCTTGTGAGGCT 1380
QY 1381 GGTGAAGAGTGTGGCGAGTGTGGTTCGCGAGGCGCCGAGGTCGAGCAGAGTTCCGT 1440
Db 1381 GGTGAAGAGTGTGGCGAGTGTGGTTCGCGAGGCGCCGAGGTCGAGCAGAGTTCCGT 1440
QY 1441 AGGTGTCGCGCGGCTTCTTCAAGACAGTTCCTGAGAGGCTGTGAGGCTTCGTAAGCT 1500
Db 1441 AGGTGTCGCGCGGCTTCTTCAAGACAGTTCCTGAGAGGCTGTGAGGCTTCGTAAGCT 1500
QY 1501 GTGCGTCTGAAATGAAGAGTATCTCCGGATAGCCCAAGATATACGACTGTTGGATGAA 1560

Db 1501 GTGCGTCTGAAATGAAGAGTATCTCCGGATAGCCCAAGATATACGACTGTTGGATGAA 1560
QY 1561 AGGCAGAGAGCGGTTGAAGGTTCTTGCAAAACCTTAGTTACCGCTACATGGGTGAGCGGC 1620
Db 1561 AGGCAGAGAGCGGTTGAAGGTTCTTGCAAAACCTTAGTTACCGCTACATGGGTGAGCGGC 1620
QY 1621 GCTAGGTGATATTCAGGAGTGCAGAAAGGCTCTCAGCGCTTCGGGTAGGACCTCAT 1680
Db 1621 GCTAGGTGATATTCAGGAGTGCAGAAAGGCTCTCAGCGCTTCGGGTAGGACCTCAT 1680
QY 1681 CGCAGCGCATCAACATAGCTTCGTAACCTAGGCTCAAGGTGATCTACGTTGACACAGAT 1740
Db 1681 CGCAGCGCATCAACATAGCTTCGTAACCTAGGCTCAAGGTGATCTACGTTGACACAGAT 1740
QY 1741 TCGCTCTTCGTGACCTATGATCCGAGAGGTGGAATAATTTTCATCAAAATTTATAAGGAG 1800
Db 1741 TCGCTCTTCGTGACCTATGATCCGAGAGGTGGAATAATTTTCATCAAAATTTATAAGGAG 1800
QY 1801 GAGCTGGGTTTCGAAATCAAGCTAGAGAGGTGTACAAACGCTTATTTTACAGAGGCT 1860
Db 1801 GAGCTGGGTTTCGAAATCAAGCTAGAGAGGTGTACAAACGCTTATTTTACAGAGGCT 1860
QY 1861 AAGAGAGGTACGCTGGCTTCTCGAGACGAGCTATAGATATTTGCGTTTCGAGGCT 1920
Db 1861 AAGAGAGGTACGCTGGCTTCTCGAGACGAGCTATAGATATTTGCGTTTCGAGGCT 1920
QY 1921 GTACGTGGCGATTCGTGTGAATCTCCAGAGAGGTTCCAGACTAAGGTTGTCGAAATAGTA 1980
Db 1921 GTACGTGGCGATTCGTGTGAATCTCCAGAGAGGTTCCAGACTAAGGTTGTCGAAATAGTA 1980
QY 1981 TTGAAGACGAGTGTGAACAAAGGCTCTCAGAGTACGTCAGAGAGATTTGTAAGAGTTG 2040
Db 1981 TTGAAGACGAGTGTGAACAAAGGCTCTCAGAGTACGTCAGAGAGATTTGTAAGAGTTG 2040
QY 2041 GAGGAGGCAAGGTTCCCATAGAGAGCTTGTAAATCTTGGAGAGCCCTTAGTAAGCGTCTT 2100
Db 2041 GAGGAGGCAAGGTTCCCATAGAGAGCTTGTAAATCTTGGAGAGCCCTTAGTAAGCGTCTT 2100
QY 2101 GAGGAGTACAAACGAGGACCAACGCTCTTCGCGAGAGAGATGCTGTACGAGGC 2160
Db 2101 GAGGAGTACAAACGAGGACCAACGCTCTTCGCGAGAGAGATGCTGTACGAGGC 2160
QY 2161 TACCGGTAAAGCCAGGACGACAGATAGGCTATGTAATAGTGAAGGCTGGCGCTATC 2220
Db 2161 TACCGGTAAAGCCAGGACGACAGATAGGCTATGTAATAGTGAAGGCTGGCGCTATC 2220
QY 2221 AGTCAAGAGCATGCCATACTTTCATGTCGATCAAGGATCTTAGCCAGATAGAGCTGACCTAC 2280
Db 2221 AGTCAAGAGCATGCCATACTTTCATGTCGATCAAGGATCTTAGCCAGATAGAGCTGACCTAC 2280
QY 2281 TATGTTGACCAACCAATCATCCCGCTGCATTGAGATACCTGGGCTACTTTGGCATCAC 2340
Db 2281 TATGTTGACCAACCAATCATCCCGCTGCATTGAGATACCTGGGCTACTTTGGCATCAC 2340
QY 2341 GAGAGAGAGCTGAAGCAAGTGCACACTGGGCGAGAGACTCTCTTCGACTTTTCTAGCCAA 2400
Db 2341 GAGAGAGAGCTGAAGCAAGTGCACACTGGGCGAGAGACTCTCTTCGACTTTTCTAGCCAA 2400
QY 2401 AAGGCAAGTAA 2412
Db 2401 AAGGCAAGTAA 2412

RESULT 3

ABX14891
ID ABX14891 standard; DNA; 2412 BP.

XX AC ABX14891;

XX DT 08-APR-2003 (first entry)

XX DE DNA encoding Pyrobolus thermostable DNA polymerase protein #2.

XX

KW Gene; ds; thermostable; DNA polymerase; DNA repair;
KW polymorphism identification.

KW polymorphism identification.

Pyrobolus fumarius.

XX	Key	Location/Qualifiers
FH		

ET	key	location
ET	CDS	1..2412

```

ET
)
/*tags a

```

```
FT /product= "Pyrobolus DNA polymerase"
```

PN US2002132243-A1.

19-SEP-2002

06-SEP-2001: 2001IS-0948369-

XX
BP 06-2117-1997, 8715-0907166

07-SEP-1999; 99US=039I340.
PR
06 SEP 0000 000000 000000 000000

20

PA (MATH/) MATHUR E

XX

XX

DR P-PSDB; ABG73162.

Novel DNA polymerase

PT and incorporating

PS Claim 1: Fig 1A-1E; 81pp; English.

CC This invention relates to a novel purified polypeptide derived from

protein of the invention is useful for catalysing the formation or

repair of a nucleic acid sequence, for comparing a sequence to a reference sequence and for identifying polymorphisms. The nucleic

acid sequence of the invention is useful for preparing cDNA from mRNA, and in an amplification process of a double-stranded DNA molecule.

The nucleotide sequence is also useful for incorporating non-natural nucleotides or intranucleotides into a DNA molecule by contacting a

polypeptide encoded by the nucleic acid with a DNA template in a PCR amplification reaction. The nucleotides which may be used for this are

inosine, 2-aminopurine or 5-methylcytosine. A nucleic acid probe

CC chromosome walking procedures to identify clones containing genomic

CC polymerases of the invention. Such methods allow the isolation of

CC polymerases of the invention have increased activity and stability at

CC sequence represents a DNA sequence encoding a pyrobolus fumarius

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Best Local Similarity 100.0%; Pred. No. 0;

[illegible]

QY I AIGACI GHAUIGTAT I CAGCGT I TTAAGAC I CTAAGC I ACAGGAACTGTTCATTACTGTTCC

DB
I AIGACIGAAAGTIGTAATCAGGGTITAGACCTACGAGGTTCTGCTATTAAGAGAGCCCTA

QY 61 CAGGTAATCATATGGGTATGCTGAGAACTGGCGAGAGGGTAGCTCTCATAGCAGGCTC

Db 61 CAGGTAATCATATGGGGTATTGCTGAGAACGGCGAGAGGGTAGTCTCATTTGACAGGTCI

Qy 121 TTTCGCCATACTTCTATGCGCTGCTTGCACCGGGCCGATCTAAGCAGGTAGCACAA

Db	121	TTTTCGCCCATACTTCTATTCGCTGCTTCGACCGGGGCGGATCCTTAAGCATGTAGCACAA	180
Qy	181	CGTATTCTGCTGCAATTCAGTAGGCGCAAGAAGCCCGATTTATAGTGTATAGAGATGACAAGAGG	240
Db	181	CGTATTCTGCTGCAATTCAGTAGGCGCAAGAAGCCCGATTTATAGTGTATAGAGATGACAAGAGG	240
Qy	241	AAGTACTTTCGGAGAGCCCTCGTAGGCTCTTAGTATTCGCACCGTCTACCCGAGGCTGTT	300
Db	241	AAGTACTTTCGGAGAGCCCTCGTAGGCTCTTAGTATTCGCACCGTCTACCCGAGGCTGTT	300
Qy	301	AGGGAGTATCGCGAACTCGTAAAGAACGTTGATGTGTTGAGGATGTTTCTTAGAGCGCGAT	360
Db	301	AGGGAGTATCGCGAACTCGTAAAGAACGTTGATGTGTTGAGGATGTTTCTTAGAGCGCGAT	360
Qy	361	ATACGCTTTCGCTATGCGCTATCTCATAGATCAAGATCTATTTCCTTTCACCTGTGACCGT	420
Db	361	ATACGCTTTCGCTATGCGCTATCTCATAGATCAAGATCTATTTCCTTTCACCTGTGACCGT	420
Qy	421	GTAGAGCTGAGCCCTCGAGAACAGATGGCTTCGGTGTGCAAGAGTATACCTGCTT	480
Db	421	GTAGAGCTGAGCCCTCGAGAACAGATGGCTTCGGTGTGCAAGAGTATACCTGCTT	480
Qy	481	AAGAGCAGCGCGAGACCTTTATGTGTAGGCTCTCGCACCAACAAAGCTTCCCGATCTT	540
Db	481	AAGAGCAGCGCGAGACCTTTATGTGTAGGCTCTCGCACCAACAAAGCTTCCCGATCTT	540
Qy	541	AGGATACTCGCTTCGATATTGAACTTTATAGCAAGCAAGGTCGCGGCTCCAGAGCGC	600
Db	541	AGGATACTCGCTTCGATATTGAACTTTATAGCAAGCAAGGTCGCGGCTCCAGAGCGC	600
Qy	601	GATCCTGTAATAGTACTGCTGAAGACTGACGTGGCGATGAGGTCCTATTTCATTGCA	660
Db	601	GATCCTGTAATAGTACTGCTGAAGACTGACGTGGCGATGAGGTCCTATTTCATTGCA	660
Qy	661	GAGGCAAGACCATCGAAAAACGATACGCGAGTTTGTAGAGTACGTGAAGAGGTATGAC	720
Db	661	GAGGCAAGACCATCGAAAAACGATACGCGAGTTTGTAGAGTACGTGAAGAGGTATGAC	720
Qy	721	CCGACATAATAGTCGGTTTATAACAAACATTCATTCGATTTGCGCTTATCTTTTGAGGCGC	780
Db	721	CCGACATAATAGTCGGTTTATAACAAACATTCATTCGATTTGCGCTTATCTTTTGAGGCGC	780
Qy	781	GCCGCGATCTTAGGCATTAAGCTTGATGTGACTAGAGAGTTGGCGCGAGCCCACT	840
Db	781	GCCGCGATCTTAGGCATTAAGCTTGATGTGACTAGAGAGTTGGCGCGAGCCCACT	840
Qy	841	AGCGTACATGGCAGCTCTCTGTCCTCGCAGGCTTAACTAGATCTGTACGACTATGCC	900
Db	841	AGCGTACATGGCAGCTCTCTGTCCTCGCAGGCTTAACTAGATCTGTACGACTATGCC	900
Qy	901	GAAGAGATGCCAGAGATCAAGATAAAGATCTCGAGAGGTTCGAGAGTATCTAGCGGTG	960
Db	901	GAAGAGATGCCAGAGATCAAGATAAAGATCTCGAGAGGTTCGAGAGTATCTAGCGGTG	960
Qy	961	ATGAAGAAGAGTGAAACCGCTTATCATCAATTTGCTGGAGATTCACAGTATTTGGGACGAC	1020
Db	961	ATGAAGAAGAGTGAAACCGCTTATCATCAATTTGCTGGAGATTCACAGTATTTGGGACGAC	1020
Qy	1021	CCGAAGAAGAGACCACTATTACTGCAATACGCGCGACGATGTCGCGCTACTTACGGC	1080
Db	1021	CCGAAGAAGAGACCACTATTACTGCAATACGCGCGACGATGTCGCGCTACTTACGGC	1080
Qy	1081	TTAGCCGGAAGAGATATTGCCGTTTGTCTATCCAGTTTCTGTAAGTAAACAGGTCCTCACTA	1140
Db	1081	TTAGCCGGAAGAGATATTGCCGTTTGTCTATCCAGTTTCTGTAAGTAAACAGGTCCTCACTA	1140
Qy	1141	GACAGGTAGTGCATGATGTTGGCTTTTCACTTGAATGTTACCTGATACGCGCGGCG	1200
Db	1141	GACAGGTAGTGCATGATGTTGGCTTTTCACTTGAATGTTACCTGATACGCGCGGCG	1200
Qy	1201	TTTAAAGATGAAGAGCTTGTCCGACCGGTTGAGCGCCACAGAGAGCTTACCGTGGC	1260
Db	1201	TTTAAAGATGAAGAGCTTGTCCGACCGGTTGAGCGCCACAGAGAGCTTACCGTGGC	1260

QY	181	CCTATTCTGTCATTTGATAGGCGCAAGAGCCCGATTTATAGTGTAGAGATGACAGAGG	240
DB	181	CCTATTCTGTCATTTGATAGGCGCAAGAGCCCGATTTATAGTGTAGAGATGACAGAGG	240
QY	241	AACTACTTCGGAGGCGCTCGTAGGCTTTACGTTATTCGACCGTCTACCGAGGCTGT	300
DB	241	AACTACTTCGGAGGCGCTCGTAGGCTTTACGTTATTCGACCGTCTACCGAGGCTGT	300
QY	301	AGGAGATATCGGAACTCGTAAGAACTGTATGTTGAGATGTTCTTAGAGCGGAT	360
DB	301	AGGAGATATCGGAACTCGTAAGAACTGTATGTTGAGATGTTCTTAGAGCGGAT	360
QY	361	ATACGCTTCGCTATGCGCTATCTCATAGATCAGATCTATTTCTTACCTGGTACCGT	420
DB	361	ATACGCTTCGCTATGCGCTATCTCATAGATCAGATCTATTTCTTACCTGGTACCGT	420
QY	421	GTAGAGGCTGAGCCCTCGAGAACAGATGGCTTCGCTGTCGACAAAGTATACCTGGTT	480
DB	421	GTAGAGGCTGAGCCCTCGAGAACAGATGGCTTCGCTGTCGACAAAGTATACCTGGTT	480
QY	481	AAGAGCAGCGCGAGGCACTTTATGTTGAGGCTCTCGCACCAACCAAGCTTCCCGATCTT	540
DB	481	AAGAGCAGCGCGAGGCACTTTATGTTGAGGCTCTCGCACCAACCAAGCTTCCCGATCTT	540
QY	541	AGGATATCTCGCTTCGATATTTAGATTTATAGCAAGGCTCGCGCTCCAGAGCGC	600
DB	541	AGGATATCTCGCTTCGATATTTAGATTTATAGCAAGGCTCGCGCTCCAGAGCGC	600
QY	601	GATCCTGTAATAGTATGCTGTAGAGCTGAGATGAGGCTGAGGCTGATTTCAATGCA	660
DB	601	GATCCTGTAATAGTATGCTGTAGAGCTGAGATGAGGCTGAGGCTGATTTCAATGCA	660
QY	661	GAGGCAAAAGACGATCGAAACCGATACGCGAGTTTGTAGAGTACGTGAAGAGGTATGAC	720
DB	661	GAGGCAAAAGACGATCGAAACCGATACGCGAGTTTGTAGAGTACGTGAAGAGGTATGAC	720
QY	721	CCGACATATAGTCTGGTTATTAACAACATCATTTTCGATGGCTTATCTTTGAGGCG	780
DB	721	CCGACATATAGTCTGGTTATTAACAACATCATTTTCGATGGCTTATCTTTGAGGCG	780
QY	781	GCCTGATATAGGCAATGATTTGATGTGACATAGAGAGTTGGCGCGAGCCACCT	840
DB	781	GCCTGATATAGGCAATGATTTGATGTGACATAGAGAGTTGGCGCGAGCCACCT	840
QY	841	AGGATATAGGCGACGCTCTCTGCTGCTGCGAGGCTTACGTAGATCTGTACGATATGCC	900
DB	841	AGGATATAGGCGACGCTCTCTGCTGCTGCGAGGCTTACGTAGATCTGTACGATATGCC	900
QY	901	GAAGAGATGCGAGATCAAGATTAAGATCTCGAGAGGTCGACAGATCTAGGCGTG	960
DB	901	GAAGAGATGCGAGATCAAGATTAAGATCTCGAGAGGTCGACAGATCTAGGCGTG	960
QY	961	ATCAAGAGAGTGAACCGCTTATCATCAATTTGTTGGAGATTCAGACTATTTGGGACGAC	1020
DB	961	ATCAAGAGAGTGAACCGCTTATCATCAATTTGTTGGAGATTCAGACTATTTGGGACGAC	1020
QY	1021	CCGAGAGAGACCACTATTAATGCAATACGCGCGGACGATGTCGCGCTACTTACGCG	1080
DB	1021	CCGAGAGAGACCACTATTAATGCAATACGCGCGGACGATGTCGCGCTACTTACGCG	1080
QY	1081	TTAGCCGAGAGATATTTGCGGTTTGTCTATCCAGTTGTCGATACAGTCTCCCACTA	1140
DB	1081	TTAGCCGAGAGATATTTGCGGTTTGTCTATCCAGTTGTCGATACAGTCTCCCACTA	1140
QY	1141	GACAGATGAGAGCTTTGCGGAAACCGGCTTACGCGGCTGAGGCTGATACCGCGGG	1200
DB	1141	GACAGATGAGAGCTTTGCGGAAACCGGCTTACGCGGCTGAGGCTGATACCGCGGG	1200
QY	1201	TTTAAAGATGAAGAGCTTTGCGGAAACCGGCTTACGCGGCTGAGGCTGATACCGGCT	1260
DB	1201	TTTAAAGATGAAGAGCTTTGCGGAAACCGGCTTACGCGGCTGAGGCTGATACCGGCT	1260
QY	1261	GCTATAGTCTTGAGCGCTTGAGAGGCTGACAGAAATATAGCGCTACTCGACTTTAGC	1320
DB	1261	GCTATAGTCTTGAGCGCTTGAGAGGCTGACAGAAATATAGCGCTACTCGACTTTAGC	1320
QY	1321	TCGATGATACCCAAACATCATGATAAAGTACAATGTTGGTCTCTGACACGCTTGTAGGCGCT	1380
DB	1321	TCGATGATACCCAAACATCATGATAAAGTACAATGTTGGTCTCTGACACGCTTGTAGGCGCT	1380
QY	1381	GGTAAAAGTGTGGCGAGTGTGGTGTCTGGAGGCGCCGAGGCTGACGACAGGCTTCCGT	1440
DB	1381	GGTAAAAGTGTGGCGAGTGTGGTGTCTGGAGGCGCCGAGGCTGACGACAGGCTTCCGT	1440
QY	1441	AGGTTCTCGCCCGCTTCTTCAAGACAGTCTTGGAGGCTGTTAGAGCTTCTGTAAGCGT	1500
DB	1441	AGGTTCTCGCCCGCTTCTTCAAGACAGTCTTGGAGGCTGTTAGAGCTTCTGTAAGCGT	1500
QY	1501	GTGCGTCTGAAATGAAGAAAGTATCTTCCGATAGCCGACCAATATCGACTTGTGGATGAA	1560
DB	1501	GTGCGTCTGAAATGAAGAAAGTATCTTCCGATAGCCGACCAATATCGACTTGTGGATGAA	1560
QY	1561	AGGCAAGAGGCTTGAAGGTTCTTGCAAAACGCTAGTTACGGCTACATGGGTTGGAGGCGC	1620
DB	1561	AGGCAAGAGGCTTGAAGGTTCTTGCAAAACGCTAGTTACGGCTACATGGGTTGGAGGCGC	1620
QY	1621	GCTAGGCTGATTTTCAGGAGTGTGCAAAAGGCTGTACGGCTTGGGCTAGGCTCTCAT	1680
DB	1621	GCTAGGCTGATTTTCAGGAGTGTGCAAAAGGCTGTACGGCTTGGGCTAGGCTCTCAT	1680
QY	1681	CGCAGCGCATCAACATAGCTCGTAAACTAGGCTCAAGGCTGATCTACGCTGACACAGAT	1740
DB	1681	CGCAGCGCATCAACATAGCTCGTAAACTAGGCTCAAGGCTGATCTACGCTGACACAGAT	1740
QY	1741	TCGCTCTTCGTCGATGATCCGAGAGGCTGAAATTTTCATCAAAATTTTAAAGGAG	1800
DB	1741	TCGCTCTTCGTCGATGATCCGAGAGGCTGAAATTTTCATCAAAATTTTAAAGGAG	1800
QY	1801	GAGCTGGGCTTCGAAATCAAGCTAGAGAGGCTGACAAACGCTTATCTTTCACAGAGCT	1860
DB	1801	GAGCTGGGCTTCGAAATCAAGCTAGAGAGGCTGACAAACGCTTATCTTTCACAGAGCT	1860
QY	1861	AAGAGAGGTAAGCTGCGCTTCTCGAGAGCGATAGATATGTCGCTTTCGAGGCT	1920
DB	1861	AAGAGAGGTAAGCTGCGCTTCTCGAGAGCGATAGATATGTCGCTTTCGAGGCT	1920
QY	1921	GTCAGTGGCTATGTTGTGAACTGCGCAAGAGGCTGACAGTATGTTGCGAAATAGTA	1980
DB	1921	GTCAGTGGCTATGTTGTGAACTGCGCAAGAGGCTGACAGTATGTTGCGAAATAGTA	1980
QY	1981	TTGAAGAGGAGTGTGAGTGAACAGGCTGTAGAGTACGTGAGAGAGTTGTGAAAGAGTTG	2040
DB	1981	TTGAAGAGGAGTGTGAGTGAACAGGCTGTAGAGTACGTGAGAGAGTTGTGAAAGAGTTG	2040
QY	2041	GAGGAGGCAAGGTTCCCATGAGAGCTTGTAAATCTGGAAGACCTTTAGTAAGGCTTT	2100
DB	2041	GAGGAGGCAAGGTTCCCATGAGAGCTTGTAAATCTGGAAGACCTTTAGTAAGGCTTT	2100
QY	2101	GAGGAGTACAAACGAGGCGCACACGCTGTCGAGCGAAGAGGCTGTCAGAGGCG	2160
DB	2101	GAGGAGTACAAACGAGGCGCACACGCTGTCGAGCGAAGAGGCTGTCAGAGGCG	2160
QY	2161	TACCGGTAAGCCCGAGGCGCAAGATAGGCTATGTAATGTAAGAGGCTGTCGCGATC	2220
DB	2161	TACCGGTAAGCCCGAGGCGCAAGATAGGCTATGTAATGTAAGAGGCTGTCGCGATC	2220
QY	2221	AGTCAAGAGGATGCGCTTACTTTCATGCTCAAGGATCTTAGCCAGATAGAGCTGACCTAC	2280
DB	2221	AGTCAAGAGGATGCGCTTACTTTCATGCTCAAGGATCTTAGCCAGATAGAGCTGACCTAC	2280
QY	2281	TATGTTGACCCAAATCATCCGCTGCAATGAGAACTGAGGCTTCTTGGGATCACC	2340
DB	2281	TATGTTGACCCAAATCATCCGCTGCAATGAGAACTGAGGCTTCTTGGGATCACC	2340
QY	2341	GAGAGAGCTGAAAGCAAGTGCACCTGCGGAGAGACTCTCTTTCGACTTTCTAGCCAG	2400

```

Db 2341 GAGAGAGCTGAGAGCAAGTGCACCTGGGCGAGAGACTCTCTCGACTTTTAGCCAG 2400
QY 2401 AAGACCAAGTAA 2412
Db 2401 AAGAGCAAGTAA 2412

RESULT 5
ID AAD35188 standard; DNA; 2412 BP.
AC AAD35188;
DT 25-JUL-2002 (first entry)
XX Pyrolobus fumaria DNA polymerase encoding DNA, PLF1831.
XX DNA polymerase; thermostable; enzyme; gene; ds.
OS Pyrolobus fumaria.
FH Key Location/Qualifiers
FT CDS 1..2412
FT /tag= a
FT /product= "Pyrolobus fumaria DNA polymerase, PLF1831"
PN WO200220735-A2.
XX 14-MAR-2002.
XX 06-SEP-2001; 2001WO-US28007.
XX 06-SEP-2000; 2000US-0656309.
XX (DIVE-) DIVERSA CORP.
XX Callen W, Mathur EJ, Short JM;
XX WPI; 2002-362247/39.
XX P-PSDB; AAE22119.
XX New thermostable polymerase useful for sequencing DNA, amplifying
XX double stranded DNA, or incorporating a non-natural nucleotide or a
XX nucleotide analog into a DNA molecule -
XX Claim 1; Fig 1P-H; 161pp; English.
XX The invention relates to thermostable DNA polymerases having high
XX temperature polymerase activity, such as those derived from
XX Pyrolobus fumaria and nucleic acid molecules encoding such polymerases.
XX Polymerases are useful for catalyzing the formation or repair of a
XX nucleic acid sequence and for modifying small molecules. They are also
XX useful for sequencing DNA molecules, for preparing cDNA from mRNA, for
XX amplifying double stranded DNA molecules and for incorporating non-
XX natural nucleotides or nucleotide analogues into a DNA molecule. The
XX present sequence is Pyrolobus fumaria DNA polymerase encoding DNA.
XX Sequence 2412 BP; 627 A; 513 C; 712 G; 560 T; 0 other;
XX Query Match 99.5%; Score 2400.8; DB 24; Length 2412;
XX Best Local Similarity 99.7%; Pred. No. 0;
XX Matches 2405; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 ATGACTGAGTGTGATTACCGGTTTGTAGCTTAGCTACGAGGTTGTTGGTAAGAGCCT 60
Db 1 ATGACTGAGTGTGATTACCGGTTTGTAGCTTAGCTACGAGGTTGTTGGTAAGAGCCT 60
QY 61 CAGGTAAATCATATGGGTTATTCCTGAGAACGCGGAGAGGTTAGTCTCATTTGACAGGTCT 120
Db 61 CAGGTAAATCATATGGGTTATTCCTGAGAACGCGGAGAGGTTAGTCTCATTTGACAGGTCT 120
QY 121 TTTCGCCCACTTCTATGCGTGTGTCACCGGCGCGGATCTCTTAAGCAGGTAGACAA 180

```

```

Db 121 TTTCGCCCACTTCTATGCGTGTGTCACCGGCGCGGATCTCTTAAGCAGGTAGACAA 180
QY 181 CGTATTCTGTCATTGAGTCCCAAGAGCCGCAATTTAGTGTGTAGAGGTGCAAGAGG 240
Db 181 CGTATTCTGTCATTGAGTCCCAAGAGCCGCAATTTAGTGTGTAGAGGTGCAAGAGG 240
QY 241 AAGTACTTTCGGGAGGCTCGTAGGGTCTTACGTATTTCGCACCGGCTTACCCGAGGCTGTT 300
Db 241 AAGTACTTTCGGGAGGCTCGTAGGGTCTTACGTATTTCGCACCGGCTTACCCGAGGCTGTT 300
QY 301 AGGGAGTATCCGGAACCTGTTAAAGAACTGTTGATGCTGTGTAGAGTGTCTAGAGGCGGAT 360
Db 301 AGGGAGTATCCGGAACCTGTTAAAGAACTGTTGATGCTGTGTAGAGTGTCTTAGAGGCGGAT 360
QY 361 ATACGCTTCGCTATGCGCTATCTCATAGTACACCATCTATTTCCTTTTCACTGTGACCGT 420
Db 361 ATACGCTTCGCTATGCGCTATCTCATAGTACACCATCTATTTCCTTTTCACTGTGACCGT 420
QY 421 GTAGAGGCTGAGCCCTCGAGAACAAAGATGGGCTTCGCTGTGCAAGAGGTATACCTGTT 480
Db 421 GTAGAGGCTGAGCCCTCGAGAACAAAGATGGGCTTCGCTGTGCAAGAGGTATACCTGTT 480
QY 481 AAGACGAGCGCGAGCCACTTTTATGCTGAGGCTCTCGCACCAACCAAGAGCTTCCGACTT 540
Db 481 AAGACGAGCGCGAGCCACTTTTATGCTGAGGCTCTCGCACCAACCAAGAGCTTCCGACTT 540
QY 541 AGGATACCTCGGCTTCGATATTCAAGTTTATAGCAAGCAAGGCTCGCGCGCTCCAGAGCGC 600
Db 541 AGGATACCTCGGCTTCGATATTCAAGTTTATAGCAAGCAAGGCTCGCGCGCTCCAGAGCGC 600
QY 601 GATCCTGTAATAGTATAGCTGTGAGACTGACCATGCGGATGAGGTGCTATTTCATTGCA 660
Db 601 GATCCTGTAATAGTATAGCTGTGAGACTGACCATGCGGATGAGGTGCTATTTCATTGCA 660
QY 661 GAGGCAAGAGCAGTCCGAAACCGATACGCGAGTTTGTAGAGTACGTGAAGAGGTATGAC 720
Db 661 GAGGCAAGAGCAGTCCGAAACCGATACGCGAGTTTGTAGAGTACGTGAAGAGGTATGAC 720
QY 721 CCGACATTAATAGTCCGTTATTAACCAATCATTTGATGCGCTTATCTTTTGAAGCGC 780
Db 721 CCGACATTAATAGTCCGTTATTAACCAATCATTTGATGCGCTTATCTTTTGAAGCGC 780
QY 781 GCCCGATCTTAGGCTAAGAGCTTGTGCTGCTAGAGAGTTGCGCGGAGCCACCACT 840
Db 781 GCCCGATCTTAGGCTAAGAGCTTGTGCTGCTAGAGAGTTGCGCGGAGCCACCACT 840
QY 841 AGCGTACATGCGCACGTCTCTGCTCGCGAGGCTTAAACGTAGATCTGTACGACTATGCC 900
Db 841 AGCGTACATGCGCACGTCTCTGCTCGCGAGGCTTAAACGTAGATCTGTACGACTATGCC 900
QY 901 GAAGAGATGCCAGAGATCAAGATAAGAGTCTCGAGAGGTCGAGAGTATCTAGGCGTG 960
Db 901 GAAGAGATGCCAGAGATCAAGATAAGAGTCTCGAGAGGTCGAGAGTATCTAGGCGTG 960
QY 961 ATGAAGAAGAGTGAACGCGTTTATCATCAATTTGTTGGAGATTTCCAGACTATTGGAGCGAC 1020
Db 961 ATGAAGAAGAGTGAACGCGTTTATCATCAATTTGTTGGAGATTTCCAGACTATTGGAGCGAC 1020
QY 1021 CCGAAGAGAGACCACTATTACTGCAATACGCGCGCGACGATGTCGCGCTACTTACGCG 1080
Db 1021 CCGAAGAGAGACCACTATTACTGCAATACGCGCGCGACGATGTCGCGCTACTTACGCG 1080
QY 1081 TTAGCCGAGAGATATTCGCTTTGCTATTCAGTTGCTGACGTAACAGGTCTCCCACTA 1140
Db 1081 TTAGCCGAGAGATATTCGCTTTGCTATTCAGTTGCTGACGTAACAGGTCTCCCACTA 1140
QY 1141 GACGAGTATGAGTGTGCTGCTTTTTCGACTTTGAAATGCTACCTGATACGCGCGG 1200
Db 1141 GACGAGTATGAGTGTGCTGCTTTTTCGACTTTTGAATGCTACCTGATACGCGCGG 1200
QY 1201 TTTAGATGAAGAGCTTTGCGCGAAGCCGCTGTGAGCGCCGAGAGACTTACCGTGGC 1260

```

1201 TTTAAGATGAAGAGCTTGTGCGCAACCGGTTGAGCGCCGACAGAGACTTACCGTGC 1260
1261 GCTATAGTCTTTCAGCGCTGAGAGCGGTGCACGAGAAATATACCGCTACTCGACTTTAGC 1320
1261 GCTATAGTCTTTCAGCGCTGAGAGCGGTGCACGAGAAATATACCGCTACTCGACTTTAGC 1320
1321 TCGATGTACCCAAACATCATGATAAAAGTACAAATGTTGGTCTGACACGCTTGTGAGGCT 1380
1321 TCGATGTACCCAAACATCATGATAAAAGTACAAATGTTGGTCTGACACGCTTGTGAGGCT 1380
1381 GGTGAAAGGTGCGGAGTGTGTTGCTGGAGCGCCCGAGAGTCAAGCACAGGTTCCGT 1440
1381 GGTGAAAGGTGCGGAGTGTGTTGCTGGAGCGCCCGAGAGTCAAGCACAGGTTCCGT 1440
1441 AGGTGTCGCCCGCGCTTCTTCAAGACAGTCTTGTGAGGCTGTTAGAGCTTCGTAGCGT 1500
1441 AGGTGTCGCCCGCGCTTCTTCAAGACAGTCTTGTGAGGCTGTTAGAGCTTCGTAGCGT 1500
1501 GTGCGTGTGCTGAAATGAAGAGTATCCTCCGATAGCCAGAAATATCGACTGTTGGATGAA 1560
1501 GTGCGTGTGCTGAAATGAAGAGTATCCTCCGATAGCCAGAAATATCGACTGTTGGATGAA 1560
1561 AGCACAAGCGTGTGAAGTCTTTCGAAACGCTAGTTACGGCTACATGGGTTGGAGCGGC 1620
1561 AGCACAAGCGTGTGAAGTCTTTCGAAACGCTAGTTACGGCTACATGGGTTGGAGCGGC 1620
1621 GCTAGTGTATTCAGGAGTGCAGAAAGCTGTACCGCTTGGGTTAGGACCTCATP 1680
1621 GCTAGTGTATTCAGGAGTGCAGAAAGCTGTACCGCTTGGGTTAGGACCTCATP 1680
1681 CGCACCGCATCAACATAGCTCGTAACTAGGCTCAAGGTGATCTACGGTGCACACAGAT 1740
1681 CGCACCGCATCAACATAGCTCGTAACTAGGCTCAAGGTGATCTACGGTGCACACAGAT 1740
1741 TCGCTCTTCGTGACCTATGATCCGAGAGAGTGTGAAATTTTCATCAAAATTTAAGGAG 1800
1741 TCGCTCTTCGTGACCTATGATCCGAGAGAGTGTGAAATTTTCATCAAAATTTAAGGAG 1800
1801 GAGTCGGGTTGCAATCAAGCTAGAGAGGTGTACAAACGCTTATCTTACAGAGGT 1860
1801 GAGTCGGGTTGCAATCAAGCTAGAGAGGTGTACAAACGCTTATCTTACAGAGGT 1860
1861 AAGAAGAGTACCTCGCTTCTCGAGGACGAGCTATAGATATTCGTTTCGAGGCT 1920
1861 AAGAAGAGTACCTCGCTTCTCGAGGACGAGCTATAGATATTCGTTTCGAGGCT 1920
1921 GTACGTGGCGATTTGGTGTGAATCTGCCAAGAGGTTTCAGACTTAAGTTTCGAAATAGTA 1980
1921 GTACGTGGCGATTTGGTGTGAATCTGCCAAGAGGTTTCAGACTTAAGTTTCGAAATAGTA 1980
1981 TTGAGACGAGTGGGTGACAGCTGCTGAGGAGGTTTCAGACTTAAGTTTCGAAATAGTA 2040
1981 TTGAGACGAGTGGGTGACAGCTGCTGAGGAGGTTTCAGACTTAAGTTTCGAAATAGTA 2040
2041 GAGGAGGCAAGGTTCCCATAGAGAGCTTGTAACTGGAAGACCTTAGTAGGCTCTT 2100
2041 GAGGAGGCAAGGTTCCCATAGAGAGCTTGTAACTGGAAGACCTTAGTAGGCTCTT 2100
2101 GAGGAGTACACACGAGGACACACACGCTGTCAGCGAAGAGGATGCTGTCAGAGGC 2160
2101 GAGGAGTACACACGAGGACACACACGCTGTCAGCGAAGAGGATGCTGTCAGAGGC 2160
2161 TACCGGTTAGCCCGGACAGATAGGTTATGTAATAGTGAAGGTTGGTGGCGTATC 2220
2161 TACCGGTTAGCCCGGACAGATAGGTTATGTAATAGTGAAGGTTGGTGGCGTATC 2220
2221 AGTCAAGAGCTGGCCATCTTTCATGCTCAAGGATCTTCAGGATAGCTGACCTAC 2280
2221 AGTCAAGAGCTGGCCATCTTTCATGCTCAAGGATCTTCAGGATAGCTGACCTAC 2280
2281 TATGTTGACCAACAAATCATCCCGTGTGATTCAGAAATCTGGGCTACTTGGCATCACC 2340
2281 TATGTTGACCAACAAATCATCCCGTGTGATTCAGAAATCTGGGCTACTTGGCATCACC 2340

QY 2341 GAGAGAGCTGGAAGAGAGTGCACCTGGGCGAGAGACTCTTTCGACTTCTAGCCAAAG 2400
Db 2341 GAGAGAGCTGGAAGAGAGTGCACCTGGGCGAGAGACTCTTTCGACTTCTAGCCAAAG 2400
QY 2401 AAGAGCAAGTAA 2412
Db 2401 AAGAGCAAGTAA 2412

RESULT 6

ABX14892
ID ABX14892 standard; DNA; 2412 BP.
XX ABX14892;
AC
XX
DT 08-APR-2003 (first entry)
XX
DE DNA encoding Pyrobolus thermostable DNA polymerase protein #3.
XX
KW Gene; ds; thermostable; DNA polymerase; DNA repair;
XX polymorphism identification.
XX Pyrobolus fumarius.
XX

Key Location/Qualifiers
CDS 1..2412
FT /*tag= a
FT /product= "Pyrobolus DNA polymerase"
XX
XX US2002132243-A1.
XX 19-SEP-2002.
XX 06-SEP-2001; 2001US-0948369.
XX 06-AUG-1997; 97US-0907166.
XX 07-SEP-1999; 99US-0391340.
XX 06-SEP-2000; 2000US-0656309.

(CALL/) CALLEN W.
PA (MATH/) MATHUR E J.
PA (SHOR/) SHORT J.
XX
XX Callen W, Mathur EJ, Short J;
XX WPI; 2003-182285/18.
XX P-PSDB; ASG76163.

Novel DNA polymerases having increased activity and stability at increased pH and temperature, useful for DNA sequencing, amplification and incorporating non-natural nucleotides or nucleotide analogs -
Claim 42; Page 52-54; 81pp; English.

This invention relates to a novel purified polypeptide derived from Pyrobolus fumari which has thermostable DNA polymerase activity. The protein of the invention is useful for catalysing the formation or repair of a nucleic acid sequence, for comparing a sequence to a reference sequence and for identifying polymorphisms. The nucleic acid sequence of the invention is useful for preparing cDNA from mRNA, and in an amplification process of a double-stranded DNA molecule. The nucleotide sequence is also useful for incorporating non-natural nucleotides or its analogues into a DNA molecule, by contacting a polypeptide encoded by the nucleic acid with a DNA template in a PCR amplification reaction. The nucleotides which may be used for this are inosine, 2-aminopurine or 5-methylcytosine. A nucleic acid probe derived from the nucleic acid sequence of the invention is useful in chromosome walking procedures to identify clones containing genomic sequences located adjacent to a sequence of the gene encoding the DNA polymerases of the invention. Such methods allow the isolation of genes which encode additional proteins from the host organisms. The

CC polymerases of the invention have increased activity and stability at
 CC increased pH and temperature, and high processivity. The present
 CC sequence represents a DNA sequence encoding a Pyrobolus fumarius
 CC thermostable DNA polymerase protein of the invention.
 XX
 SQ Sequence 2412 BP; 627 A; 513 C; 712 G; 560 T; 0 other;

Query Match 99.5%; Score 2400.8; DB 25; Length 2412;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2405; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY	1	ATGACTGAAGTGTGTAATTCACCGTCTTTAGACTCTAGCTACGAGGTGTGTTGTAAGAGCGCT	60
DB	1	ATGACTGAAGTGTGTAATTCACCGTCTTTAGACTCTAGCTACGAGGTGTGTTGTAAGAGCGCT	60
QY	61	CAGGTAATCATATAGGGGTATTCCTGAGAAACCGCGAGAGGTAGTCTCATATGACAGCTCT	120
DB	61	CAGGTAATCATATAGGGGTATTCCTGAGAAACCGCGAGAGGTAGTCTCATATGACAGCTCT	120
QY	121	TTTCGCCCATCTCTATGCGCTCTTGCACCGCGCGCCGATCCTAGACAGGTAGACACAA	180
DB	121	TTTCGCCCATCTCTATGCGCTCTTGCACCGCGCGCCGATCCTAGACAGGTAGACACAA	180
QY	181	CGTATTCGTGCTAGTAGGAGCCAAAGAGCCCGGATATATAGGTGTAGAGGTATGCAAGAGG	240
DB	181	CGTATTCGTGCTAGTAGGAGCCAAAGAGCCCGGATATATAGGTGTAGAGGTATGCAAGAGG	240
QY	241	AAGTACTTCGGAGGCGCTCTAGAGGTCTTAGCTATTCGACCGGTCTACCGAGGCTGTT	300
DB	241	AAGTACTTCGGAGGCGCTCTAGAGGTCTTAGCTATTCGACCGGTCTACCGAGGCTGTT	300
QY	301	AGGAGATTCGCGAACTCGTAAAGAAACGTTGATGGTGTGAGGATGTTCTAGAGGCGGAT	360
DB	301	AGGAGATTCGCGAACTCGTAAAGAAACGTTGATGGTGTGAGGATGTTCTAGAGGCGGAT	360
QY	361	ATACGCTTCGCTATGCGCTATCTCATAGATCAGCATCTATTTCTTTCACTGTTACCGT	420
DB	361	ATACGCTTCGCTATGCGCTATCTCATAGATCAGCATCTATTTCTTTCACTGTTACCGT	420
QY	421	GTAGAGGCTGAGCCCTCGAGAAACGATGAGGCTTCGCTGTCGACAAAGGTATACCTGTT	480
DB	421	GTAGAGGCTGAGCCCTCGAGAAACGATGAGGCTTCGCTGTCGACAAAGGTATACCTGTT	480
QY	481	AAGACAGCCCGGAGCCACTTTATGGTGAAGCTCTCGACACAAACCAAGCTTCGCGATCTT	540
DB	481	AAGACAGCCCGGAGCCACTTTATGGTGAAGCTCTCGACACAAACCAAGCTTCGCGATCTT	540
QY	541	AGGATACCTCGCTTCGATTTGAGCTTTATAGCAAGCAAGGTCGCGCTCCAGAGCCG	600
DB	541	AGGATACCTCGCTTCGATTTGAGCTTTATAGCAAGCAAGGTCGCGCTCCAGAGCCG	600
QY	601	GATCTCTGTAATAGTGATAGCTGTGAAGACTCAGCATGCGATGAGGTGCTATTCATTGCA	660
DB	601	GATCTCTGTAATAGTGATAGCTGTGAAGACTCAGCATGCGATGAGGTGCTATTCATTGCA	660
QY	661	GAGGCAACAGCATCGAAACCGATACGAGTTGTAGAGTACGTGAAGAGGTATGAC	720
DB	661	GAGGCAACAGCATCGAAACCGATACGAGTTGTAGAGTACGTGAAGAGGTATGAC	720
QY	721	CCGACATATAGTCCGCTTATAACAAATCATTTTCGATTGCGCTTATCTTTGAGGCGC	780
DB	721	CCGACATATAGTCCGCTTATAACAAATCATTTTCGATTGCGCTTATCTTTGAGGCGC	780
QY	781	GCCCGATCTTAGGCATTAAGCTTGATGTGATGATGAGAGTTGGCGCGAGCCACCAT	840
DB	781	GCCCGATCTTAGGCATTAAGCTTGATGTGATGATGAGAGTTGGCGCGAGCCACCAT	840
QY	841	ACGCTACATGGCAGCTCTCTGCTCCGCGAGGCTTAACGTAGATCTGTACGACTATGCC	900
DB	841	ACGCTACATGGCAGCTCTCTGCTCCGCGAGGCTTAACGTAGATCTGTACGACTATGCC	900
QY	901	GAAGAGATCCAGAGATCAAGATAAGAGTCTCGAGGAGGTTCGACAGGTATCTAGGCGTG	960
DB	901	GAAGAGATCCAGAGATCAAGATAAGAGTCTCGAGGAGGTTCGACAGGTATCTAGGCGTG	960

DB	901	GAAGAGATCCAGAGATCAAGATAAGAGTCTCGAGGAGGTTCGACAGGTATCTAGGCGTG	960
QY	961	ATGAGAGAGAGTGAACCGGTATATCATCAATTTGTGGAGATTCGACACTATTCGGACGAC	1020
DB	961	ATGAGAGAGAGTGAACCGGTATATCATCAATTTGTGGAGATTCGACACTATTCGGACGAC	1020
QY	1021	CCGAGAGAGAGACCACTATTACTGCAATACGCGCGGACGATCTCCGCGCTACTTACGGC	1080
DB	1021	CCGAGAGAGAGACCACTATTACTGCAATACGCGCGGACGATCTCCGCGCTACTTACGGC	1080
QY	1081	TTAGCCGAGAAATATTCGCGTCTTATCCAGTTGTCTGATGCTGATACGAGCTCTCCCACTA	1140
DB	1081	TTAGCCGAGAAATATTCGCGTCTTATCCAGTTGTCTGATGCTGATACGAGCTCTCCCACTA	1140
QY	1141	GACCAAGTGTGCGATGAGTGTGGCTTTTCGACTTGAATGGTACCTGATACCGCGGCG	1200
DB	1141	GACCAAGTGTGCGATGAGTGTGGCTTTTCGACTTGAATGGTACCTGATACCGCGGCG	1200
QY	1201	TTTAAGATGAAGAGCTTGTGCCGAACCGGTGTGAGCCGCCAGAGAGACTTACCGTGGC	1260
DB	1201	TTTAAGATGAAGAGCTTGTGCCGAACCGGTGTGAGCCGCCAGAGAGACTTACCGTGGC	1260
QY	1261	GCTATAGTCTTTAGCCGCTTTGAGAGCGGTGCAAGAAATATAGCGTATCTGAGCTTTAGC	1320
DB	1261	GCTATAGTCTTTAGCCGCTTTGAGAGCGGTGCAAGAAATATAGCGTATCTGAGCTTTAGC	1320
QY	1321	TCGATGTACCCCAACATCATGATAAAGTACAAATTTGCTCTGACAGCTTTGAGGCGCT	1380
DB	1321	TCGATGTACCCCAACATCATGATAAAGTACAAATTTGCTCTGACAGCTTTGAGGCGCT	1380
QY	1381	GGTGAAGAGTGTGGCGAGTGTGGTGTGGAGGCGCCCGAGAGTCAAGACAGGTTCCGT	1440
DB	1381	GGTGAAGAGTGTGGCGAGTGTGGTGTGGAGGCGCCCGAGAGTCAAGACAGGTTCCGT	1440
QY	1441	AGGTGTCCGCGCGCTTCTTCAAGACAGCTTCTGAGAGGCTTTGAGAGCTTTGAGGCGCT	1500
DB	1441	AGGTGTCCGCGCGCTTCTTCAAGACAGCTTCTGAGAGGCTTTGAGAGCTTTGAGGCGCT	1500
QY	1501	GTGCGTGTGTAATGAAGAGTATCTCTCCGATAGCCAGAAATATCGACTTGTGGATGAA	1560
DB	1501	GTGCGTGTGTAATGAAGAGTATCTCTCCGATAGCCAGAAATATCGACTTGTGGATGAA	1560
QY	1561	AGGCAAGAGGCTTGAAGGTTCTTGCAAAACGCTAGTTTACCGCTACATGGTTCGAGCGC	1620
DB	1561	AGGCAAGAGGCTTGAAGGTTCTTGCAAAACGCTAGTTTACCGCTACATGGTTCGAGCGC	1620
QY	1621	GCTAGTGTGTATTCAGGAGTGTGCGCAAGGCTCTCAGCGCTTGGGCTAGGCACTCATTA	1680
DB	1621	GCTAGTGTGTATTCAGGAGTGTGCGCAAGGCTCTCAGCGCTTGGGCTAGGCACTCATTA	1680
QY	1681	CGCACCGCATCAACATAGTCTGTAAGCTCAAGGCTTCAAGGTGATCTACCGTGACACAGAT	1740
DB	1681	CGCACCGCATCAACATAGTCTGTAAGCTCAAGGCTTCAAGGTGATCTACCGTGACACAGAT	1740
QY	1741	TCGCTCTTCGTGACCTATGATCCGAGAGGTGGAABATTTTCATCAAAATTTAAGAGGAG	1800
DB	1741	TCGCTCTTCGTGACCTATGATCCGAGAGGTGGAABATTTTCATCAAAATTTAAGAGGAG	1800
QY	1801	GAGCTGGGGTTCGAAATCAAGCTAGAGAGGTGTACAAACGCTTATTTTACAGAGGCT	1860
DB	1801	GAGCTGGGGTTCGAAATCAAGCTAGAGAGGTGTACAAACGCTTATTTTACAGAGGCT	1860
QY	1861	AGAAGAGGTACGCTGCGCTTCTCAGGACGAGTATAGATATTTGTCGGTTTCGAGGCT	1920
DB	1861	AGAAGAGGTACGCTGCGCTTCTCAGGACGAGTATAGATATTTGTCGGTTTCGAGGCT	1920
QY	1921	GTACGTCGCGATTTGTTGTAAGTCTCCCAAGAGGTTCAGACTAAGGTTCGAAATAGTA	1980
DB	1921	GTACGTCGCGATTTGTTGTAAGTCTCCCAAGAGGTTCAGACTAAGGTTCGAAATAGTA	1980
QY	1981	TTGAAGACGAGTGTGACCAAGGCTGTAGAGTACGCTCAGGAGAGATTTGTAAGAGGTTG	2040
DB	1981	TTGAAGACGAGTGTGACCAAGGCTGTAGAGTACGCTCAGGAGAGATTTGTAAGAGGTTG	2040

QY 2041 GAGGAGGCGAGGTTCCCATAGAGAGCTTGTATCTGAAAGACCCCTAGTAAGCGTCTT 2100
 Db |||||||
 QY 2041 GAGGAGGCGCAAGTTCCCATAGAGAGCTTGTATCTGAAAGACCCCTAGTAAGCGTCTT 2100
 Db |||||||
 QY 2101 GAGGAGTACACACGAGGACCCACACGTCGTTGAGGAGAGAGATCTCTCAGCAGGC 2160
 Db |||||||
 QY 2101 GAGGAGTACACACGAGGACCCACACGTCGTTGAGGAGAGAGATCTCTCAGCAGGC 2160
 Db |||||||
 QY 2161 TACCGGTAAGCCAGGCGACAGATAGGATATGTAATAGTGAAGGGTGGTGGCGGTATC 2220
 Db |||||||
 QY 2161 TACCGGTAAGCCAGGCGACAGATAGGATATGTAATAGTGAAGGGTGGTGGCGGTATC 2220
 Db |||||||
 QY 2221 AGTCAAGAGATGCGGCTACTTCTATGTCAGGATCCCTAGCCAGATAGAGCGTACCTAC 2280
 Db |||||||
 QY 2221 AGTCAAGAGATGCGGCTACTTCTATGTCAGGATCCCTAGCCAGATAGAGCGTACCTAC 2280
 Db |||||||
 QY 2281 TATGTTGACCAACCAATCATCCGCTGCATTGAGATACCTGGCTACTTTGGCATCAC 2340
 Db |||||||
 QY 2281 TATGTTGACCAACCAATCATCCGCTGCATTGAGATACCTGGCTACTTTGGCATCAC 2340
 Db |||||||
 QY 2341 GAGAGAGAGCTGAAGAGAGTGCACCTGGGAGAGAGCTCTCTCGACTTTCTAGCCCAAG 2400
 Db |||||||
 QY 2341 GAGAGAGAGCTGAAGAGAGTGCACCTGGGAGAGAGCTCTCTCGACTTTCTAGCCCAAG 2400
 Db |||||||
 QY 2401 AAGACCAAGTAA 2412
 Db |||||||
 QY 2401 AAGACCAAGTAA 2412
 Db |||||||

RESULT 7

AAD35189

ID AAD35189 standard; DNA; 2405 BP.

AC AC

XX AAD35189;

XX 25-JUL-2002 (first entry)

XX DE

XX Pyrolobus fumaria DNA polymerase consensus DNA sequence.

XX KW

XX DNA polymerase; thermostable; enzyme; ds.

XX OS

XX Pyrolobus fumaria.

XX PN

XX WO200220735-A2.

XX FX

XX 14-MAR-2002.

XX PF

XX 06-SEP-2001; 2001WO-US28007.

XX PR

XX 06-SEP-2000; 2000US-0656309.

XX PA

XX (DIVE-) DIVERSA CORP.

XX PI

XX Callen W, Mathur EJ, Short JW;

XX DR

XX WPT; 2002-362247/39.

XX PT

XX New thermostable polymerase useful for sequencing DNA, amplifying
 XX double stranded DNA, or incorporating a non-natural nucleotide or a
 XX nucleotide analog into a DNA molecule -

XX PS

XX Disclosure; Fig 1F-H; 161pp; English.

XX CC

XX The invention relates to thermostable DNA polymerases having high
 XX temperature polymerase activity, such as those derived from
 XX Pyrolobus fumaria and nucleic acid molecules encoding such polymerases.
 XX Polymerases are useful for catalysing the formation or repair of a
 XX nucleic acid sequence and for modifying small molecules. They are also
 XX useful for sequencing DNA molecules, for preparing cDNA from mRNA, for
 XX amplifying double stranded DNA molecules and for incorporating non-
 XX natural nucleotides or nucleotide analogues into a DNA molecule. The
 XX present sequence is Pyrolobus fumaria DNA polymerase consensus DNA

CC sequence.

XX SQ Sequence 2405 BP; 627 A; 512 C; 706 G; 560 T; 0 other;
 Query Match 96.9%; Score 2338; DB 24; Length 2405;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2405; Conservative 0; Mismatches 0; Indels 7; Gaps 6;
 QY 1 ATGACTGAAGTCTTATTCACGGTTTATGACTCTAGCTAGCAGGTTCTGTGTAAGAGCCT 60
 Db |||||||
 QY 1 ATGACTGAAGTCTTATTCACGGTTTATGACTCTAGCTAGCAGGTTCTGTGTAAGAGCCT 60
 Db |||||||
 QY 61 CAGGTAAATCATATGCGGTATGCTGAGAACCGGCGAGAGGTAGTCTCTCAATGACGTTCT 120
 Db |||||||
 QY 61 CAGGTAAATCATATGCGGTATGCTGAGAACCGGCGAGAGGTAGTCTCTCAATGACGTTCT 120
 Db |||||||
 QY 121 TTTGCGCCATPACTTCTATGCGCTCTTATGCGCTGCTTGCACCGGCGCGGATCTTAAGCAGGTAGACAA 180
 Db |||||||
 QY 121 TTTGCGCCATPACTTCTATGCGCTGCTTGCACCGGCGCGGATCTTAAGCAGGTAGACAA 180
 Db |||||||
 QY 181 CGTATTGCTGCTTATGAGTAGGCGCAAGAGAGCCCGATTTATAGGTAGAGATGACAGAGG 240
 Db |||||||
 QY 181 CGTATTGCTGCTTATGAGTAGGCGCAAGAGAGCCCGATTTATAGGTAGAGATGACAGAGG 240
 Db |||||||
 QY 241 AAGTACTTCGGGAGGCTCTGAGGCTTACGTAATTCGACCGTGTCTACCCGAGGCTGTT 300
 Db |||||||
 QY 241 AAGTACTTCGGGAGGCTCTGAGGCTTACGTAATTCGACCGTGTCTACCCGAGGCTGTT 300
 Db |||||||
 QY 301 AGGGAGTATCGCGAAGTCTGTAAGAGACCTGATGAGTGTGAGGATGTTCTAGAGGCGGAT 360
 Db |||||||
 QY 301 AGGGAGTATCGCGAAGTCTGTAAGAGACCTGATGAGTGTGAGGATGTTCTAGAGGCGGAT 360
 Db |||||||
 QY 361 ATACGCTTCGCTATGCGCTATCTCATAGATCAGATCTATTTCTTTCCACCTGATACCGT 420
 Db |||||||
 QY 361 ATACGCTTCGCTATGCGCTATCTCATAGATCAGATCTATTTCTTTCCACCTGATACCGT 420
 Db |||||||
 QY 421 GTAGAGCTGAGCCCTCGAGAACAGATGGGCTTCCGTGTGACAAAGTATACCTGGTT 480
 Db |||||||
 QY 421 GTAGAGCTGAGCCCTCGAGAACAGATGGGCTTCCGTGTGACAAAGTATACCTGGTT 480
 Db |||||||
 QY 481 AAGAGCAGCCCGGAGCCACTTTATGCTGAGGCTCTCGACCAACAGCTTCCCGATCTT 540
 Db |||||||
 QY 481 AAGAGCAGCCCGGAGCCACTTTATGCTGAGGCTCTCGACCAACAGCTTCCCGATCTT 540
 Db |||||||
 QY 541 AGGATATCTCGGTTTCGATATTGAAGTTTATAGCAAGAGGTCGCGCGTCCAGAGCGC 600
 Db |||||||
 QY 541 AGGATATCTCGGTTTCGATATTGAAGTTTATAGCAAGAGGTCGCGCGTCCAGAGCGC 600
 Db |||||||
 QY 601 GATCCTGTATAGTATGATGCTGTGAAGACTGACGATGGCGATGAGTGTATTCATTGCA 660
 Db |||||||
 QY 601 GATCCTGTATAGTATGATGCTGTGAAGACTGACGATGGCGATGAGTGTATTCATTGCA 660
 Db |||||||
 QY 661 GAGGCAAAAGACGATCGAAACCGATACCGGAGTTGTAGAGTACGCTGAAGAGGTATGAC 720
 Db |||||||
 QY 661 GAGGCAAAAGACGATCGAAACCGATACCGGAGTTGTAGAGTACGCTGAAGAGGTATGAC 720
 Db |||||||
 QY 721 CCCGACATATAGTTCGGTTTATACCAATCATTTTCGATTTGGCTTATCTTTTGGCGCG 780
 Db |||||||
 QY 721 CCCGACATATAGTTCGGTTTATACCAATCATTTTCGATTTGGCTTATCTTTTGGCGCG 780
 Db |||||||
 QY 781 GCCCGATCCTAGGCGATAAAGCTTGAATGACTAGAGAGTTGGCGCGACCCCACT 840
 Db |||||||
 QY 781 GCCCGATCCTAGGCGATAAAGCTTGAATGACTAGAGAGTTGGCGCGACCCCACT 840
 Db |||||||
 QY 841 AGCGTACATGCGGCGCTCTCTCTCCCTGCGAGCTTAAGCTAGATCTGACACTATGCC 900
 Db |||||||
 QY 841 AGCGTACATGCGGCGCTCTCTCTCCCTGCGAGCTTAAGCTAGATCTGACACTATGCC 900
 Db |||||||
 QY 901 GAAGAGATCCAGAGATCAAGATAAAGAGTCTCGAGGAGGTCGAGAGTATCTAGCGGTG 960
 Db |||||||
 QY 901 GAAGAGATCCAGAGATCAAGATAAAGAGTCTCGAGGAGGTCGAGAGTATCTAGCGGTG 960
 Db |||||||
 QY 961 ATGAAGAGAGTGAACCGCTTATCATCAATTTGGTGGGAGATTCCAGACTATTGGGAGCAG 1020

961 ATGAGAGAGAGTGAACGGTATCATCAATTGGTGGAGATTCAGACTATTGGGACGAC 1020
 1021 CCGAAGAGAGACCACTATTACTCAATACGCGCGGACGATGTCGCGCTACTTACGCGC 1080
 1022 CCGAAGAGAGACCACTATTACTCAATACGCGCGGACGATGTCGCGCTACTTACGCGC 1080
 1081 TTAGCCGAGAGAGATTTGCGGTTGCTTATCCAGTTGTCGTAATACAGGTTCCCACTA 1140
 1082 TTAGCCGAGAGAGATTTGCGGTTGCTTATCCAGTTGTCGTAATACAGGTTCCCACTA 1140
 1141 GACCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 1142 GACCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 1201 TTTAAGATGAAAGAGCTTGTGCCGACCGGTTGAGCGCCACAGAGAGACTTACCCTGCG 1260
 1261 GCTATAGTCTTGGAGCGGTTGAGAGCGGTCGACGAGGATGATGATGATGATGATGATGAT 1320
 1262 GCTATAGTCTTGGAGCGGTTGAGAGCGGTCGACGAGGATGATGATGATGATGATGATGAT 1320
 1321 TCGATGTACCCCAACATCATGATTAAGTACAATGTTGCTGACACGCTTGTGAGGCTT 1380
 1322 TCGATGTACCCCAACATCATGATTAAGTACAATGTTGCTGACACGCTTGTGAGGCTT 1380
 1381 GGTGAAGAGTGTGCGAGTGTGTTGCTGGAGGCGCCCGAGGTCGACGACAGGTTCCGT 1440
 1382 GGTGAAGAGTGTGCGAGTGTGTTGCTGGAGGCGCCCGAGGTCGACGACAGGTTCCGT 1440
 1383 GGTG- AAGAGTGTGCGAGTGTGTTGCTGGAGGCGCCCGAGGTCGACGACAGGTTCCGT 1439
 1441 AGGTGTGCGCGCGCTTCTCAAGACAGTTCCTGAGAGGCTGTTAGAGCTTCTGAGGCT 1500
 1442 AGGTGTGCGCGCGCTTCTCAAGACAGTTCCTGAGAGGCTGTTAGAGCTTCTGAGGCT 1500
 1443 AGGTGTGCGCGCGCTTCTCAAGACAGTTCCTGAGAGGCTGTTAGAGCTTCTGAGGCT 1499
 1501 GTGCGTGTGAAATGAAGAGTATCTCCGAGATGACCCAGAGATGATGATGATGATGATGAT 1560
 1502 GTGCGTGTGAAATGAAGAGTATCTCCGAGATGACCCAGAGATGATGATGATGATGATGAT 1560
 1503 GTGCGTGTGAAATGAAGAGTATCTCCGAGATGACCCAGAGATGATGATGATGATGATGAT 1559
 1561 AGGCAGAGAGGCTTGAAGGTTCTTCAAGACGCTGATGTTACGCTGATGATGATGATGATGAT 1620
 1562 AGGCAGAGAGGCTTGAAGGTTCTTCAAGACGCTGATGTTACGCTGATGATGATGATGATGAT 1620
 1563 AGGCAGAGAGGCTTGAAGGTTCTTCAAGACGCTGATGTTACGCTGATGATGATGATGATGAT 1619
 1621 GCTAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
 1622 GCTAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
 1681 CGCAGCGGCTATCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
 1682 CGCAGCGGCTATCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
 1679 CGCAGCGGCTATCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1738
 1741 TCGCTCTTCGTGACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800
 1739 TCGCTCTTCGTGACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1795
 1801 GAGCTGGGTTGGAATCAAGCTAGAGAGGTTGATGATGATGATGATGATGATGATGATGATGAT 1860
 1796 GAGCTGGGTTGGAATCAAGCTAGAGAGGTTGATGATGATGATGATGATGATGATGATGATGAT 1854
 1861 AAGAGAGGCTAGCTGGCTTCTCGAGGACGAGTATGATGATGATGATGATGATGATGATGATGAT 1920
 1855 AAGAGAGGCTAGCTGGCTTCTCGAGGACGAGTATGATGATGATGATGATGATGATGATGATGAT 1914
 1921 GTAAGTGGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
 1915 GTACGTGGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1974
 1981 TTGAAGAGGCTAGCTGGCTTCTCGAGGACGAGTATGATGATGATGATGATGATGATGATGATGAT 2040
 1975 TTGAAGAGGCTAGCTGGCTTCTCGAGGACGAGTATGATGATGATGATGATGATGATGATGATGAT 2033
 2041 GAGGAGGCGAAGGTTTCCATAGAGAGCTTGTATCTGGAAGACCTTGTAGTAAAGCTCTT 2100

2034 GAGGAGGCGAAGGTTCCCATAGAGAGCTTGTATCTGGAAGACCCCTTAGTAAAGCTCTT 2093
 2101 GAGGAGTACACAAACGAGGACCCACACGCTGCTTGCAGCGAAGAGGATGCTGTACGAGC 2160
 2094 GAGGAGTACACAAACGAGGACCCACACGCTGCTTGCAGCGAAGAGGATGCTGTACGAGC 2153
 2161 TACCGGTAAGCCGAGGACGAGATAGGGTATGTAATAGTGAAGGTTGGTGGCGGTATC 2220
 2154 TACCGGTAAGCCGAGGACGAGATAGGGTATGTAATAGTGAAGGTTGGTGGCGGTATC 2213
 2221 AGTCAAAGACGATGGCCATCTTCTGTCAGGATCCTAGCAGATAGAGCTGACCTAC 2280
 2214 AGTCAAAGACGATGGCCATCTTCTGTCAGGATCCTAGCAGATAGAGCTGACCTAC 2273
 2281 TATGTTGACCAACAAATCATCCGCTGCTGATGAGATGATGAGTACTGGCTTACCTTGGCATCACC 2340
 2274 TATGTTGACCAACAAATCATCCGCTGCTGATGAGATGATGAGTACTGGCTTACCTTGGCATCACC 2333
 2341 GAGGAGAGGCTGAAGCAAGTGCACCTGGCGAGAGACTCTCTGACTTCTAGCCAG 2400
 2334 GAGGAGAGGCTGAAGCAAGTGCACCTGGCGAGAGACTCTCTGACTTCTAGCCAG 2393
 2401 ARGAGCAAGTAA 2412
 2394 AAGAGCAAGTAA 2405

RESULT 8
 AAT16274
 ID AAT16274 standard; DNA; 3068 BP.
 AC AAT16274;
 XX
 DT 12-JUL-1996 (first entry)
 XX
 DE Pyrodictium occultum 3068 bp DNA polymerase gene.
 XX
 KW DNA pol; base pair; thermally stable; exonuclease activity; ds.
 XX
 OS Pyrodictium occultum DSM2709.
 XX
 FH Key Location/Qualifiers
 FT CDS 435..3179
 FT /*tag= a
 XX /product= DNA_polymerase
 PN JP07327684-A.
 XX
 PD 19-DEC-1995.
 XX
 PF 09-JUN-1994; 94JP-0150591.
 XX
 PR 09-JUN-1994; 94JP-0150591.
 XX
 PA (TAKI) TAKARA SHUZO CO LTD.
 XX
 DR WPI; 1996-072342/08.
 DR P-PSDB; AAR92523.
 XX
 PT DNA encoding DNA polymerase - useful for prodn. of thermally stable
 PT enzyme
 PS Claim 1; Page 18-22; 23pp; Japanese.
 CC AAT16273 and AAT16274 encode a 914 amino acid and an 803 amino acid DNA
 CC polymerase, respectively. The genes are derived from Pyrodictium
 CC occultum and produce thermostable DNA polymerase. Also disclosed are
 CC DNA polymerase genes which hybridize to the above genes.
 SQ Sequence 3068 BP; 730 A; 745 C; 980 G; 613 T; 0 other;

Query Match 44.4%; Score 1071.2; DB 17; Length 3068;
 Best Local Similarity 66.5%; Pred. No. 0;

Matches 1602; Conservative 0; Mismatches 788; Indels 18; Gaps 4;

```
QY 11 TTGATTACGGTTTGTAGACTCTAGCTACGAGGTTCTTGGTAAAGAGCCTCAGTAATCA 70
Db 289 TAGAGTTCTGCTGCTAGACTCTAGCTACGAGATCTGGGAAGAGCCGGTAGTAATCC 348
QY 71 TATGGGGTATTTGCTGAGAACGGCGAGAGGGTAGTCTCATTTGACAGGTTCTTTTCGCCAT 130
Db 349 TCTGGGGGTAACCGCTTGACGGTAACAGTGTGCTCTTACAGCAACCGCTTCCGCCCT 408
QY 131 ACTTCTATGCGTCTTGCACCGCGCGCCGATCCTA-----AGCAGGTAGCACACGTA 184
Db 409 ACTTCTACGCCCTCATAGCCCGGGCTATGAGGATATGTTGGAGGATAGAGCTTCCA 468
QY 185 TTGCTGCATGTAGTGGCCAAAGAGCCCGATTATAGGTGTAGAGATGACAAGAGAAAT 244
Db 469 TAAGGAGGCTTATGTGTGTCAGAGTCCGATAATAGATGCCAAGCCTCTTTGATAAGAGT 528
QY 245 ACTTCGGGAGGCTCGTAGGGTCTTACGTTATTCGACCGTCTACCCGAGGCTGTAGGG 304
Db 529 ACTTCGGCAGGCCCGGTAAGCGGTGAAGATTACCACTATGATACCCGAGTCTGTAGAC 588
QY 305 AGTATCGGAACCTCGTAAGAACGTTGATGTGTGAGGATGTTCTAGAGGGGGATATAC 364
Db 589 ACTACCGGAGCGGTGAAGAGATAGAGGGTGTGAGGACTCCCTCGAGGCGAGTATTA 648
QY 365 GCTTCGCTATGGCTATCTCATAGATACAGATCTATTTCTTTCACCTGGTACCGGTAG 424
Db 649 GGTTTGCAATGATGATCTGATAGATAGAGGCTCTACCCGTTACCGGTTACCGGATCC 708
QY 425 AGGCTGAGCCCCCTCGAGAACAGATGGGCTCCGTCGACCAACCAAGCTTCCCGATCTAGGA 544
Db 709 CCGTAGAGGATCGGCGCCGCAATCCAGGCTTCCGTTGACCGTGTACAAAGGTTGCTG 768
QY 485 CGAGCGCGGAGCCACTTTATGTGAGGCTCTCGACCAACCAAGCTTCCCGATCTAGGA 544
Db 769 GCGACCGGAGCCCT-----AGCGGATATAACCGGATCGACCTTCCCGGATAGGC 822
QY 545 TACTCGCGTTTCATATTGAGTTTATAGCAACAGAGGTGCGCGCTCCAGAGCGCGATC 604
Db 823 TGGTAGCTTTTGATATAGAGGTGTATAGCAGAGGGGAGCCCTAACCTTCGAAGGATC 882
QY 605 CTGTAAATAGTGTAGCTGTGAAGACTGACGATGGCGATGAGTGTCTATTCTCCGAGG 664
Db 883 CAGTGTATATAGTGTGCTGTAGGGACAGCAGGCGAGGAGGAGGCTCATAGAGCTGAAG 942
QY 665 GCAAGAGCATGCAAAACCGATACCGGAGTTTGTAGTACGTGAAGAGGTATGACCCCG 724
Db 943 GCCATGACGACAGGAGGGTCTGAGGGAGTTCTGATAGTACGTGAGAGCCTTCGACCCCG 1002
QY 725 ACATAATAGTCGGTTATTAACACATCATTTGATGTCCTTATCTTTTGGAGCGGCC 784
Db 1003 ACATAATAGTGGGCTATTAACAGTACCACTTCGACTGGCCCTTACCTAATGAGCGGCC 1062
QY 785 GCATCTAGGCATTAAGCTTGATGTGACTAGAAGAGTTGGCGCGGAGCCACCACTAGCG 844
Db 1063 GTAGCTCGGATTAAGCTTCAGCTTACACGTTACACCGGTGAGGAGGAGCCCAACCG 1122
QY 845 TACATGGCACGCTCTCTGCTCCCTGGCAGGCTTAACGTAGATCTGTACGACTATGCCGAG 904
Db 1123 TCTACGGCCACGCTCTCGGTGAGGATAGGCTGACGTTGACCTCTACGACTATGCCGAGG 1182
QY 905 AGATGCCAGAGTCAAGTAAAGTCTCGAGGAGGTCGAGATATCTAGGCTGTATGA 964
Db 1183 AGATGCCGAGATTAAGTAAAGTACGCTTGAAGAGGAGTACCGGAGTACCTAGGCGTTATGA 1242
QY 965 AGAAGGTGAACGCTTATCATCAATTTGTCGGAGATTCAGACTATTTGGACACCGCA 1024
Db 1243 AGAAGAGCGCGTGTGATATATAGAGTGTGGAGATACCCGAGTACTGGGATGACAGA 1302
QY 1025 AGAAGAGACCACTATTACTGCAATACCGCGCGAGATGTCCGCTTACTTACGGCTTAG 1084
Db 1303 AGAAGAGCGAGCTGTAGAGCGCTACCGCTTCGAGGATGTAGGAGGCTACCTACGGCTCG 1362
```

```
QY 1085 CCGAAGAGATATTGCGGTTTGTCTTACCTACGTTAGTGTCTACGTAACAGGCTCTCCACTAGACC 1144
Db 1363 CGEAAAAGATGCTTACCGTTCCGCCATACAGCTCTCACTGTGTACGGGTGCTCTCTCGACC 1422
QY 1145 AGTAGGTGCGATGAGTGTGGCTTTGCACTTGAATGTAATGTAATGTAATGTAATGTAATGTA 1204
Db 1423 AGTAGGTGCTATGGCGCTAGGCTTCCGCCCTAGAGTGTGATCTCATGCTGTCAGCCTACG 1482
QY 1205 AGATGAAGAGCTTGTGCGCAACCGCTTNGCGCCCGAGAGAGACTTACCGTGGCGCTA 1264
Db 1483 ATATGAAGAGCTTGTGCGCAACCGCTTNGCGCCCGAGAGAGACTTACAGGCTGAG 1542
QY 1265 TAGTCTTGTAGCCCTTGTAGAGCGGTGCAAGAGATATAGCCGTAAGTCTGAGCTTTAGCTCGA 1324
Db 1543 TAGTGTAAAGCCTCTCAAGGAGTCCATGAGATGTGTGCTGCTCGATTTAGTTCCA 1602
QY 1325 TGTACCCAAACATCATGATTAAGTACAAATGTTGCTCTGACACGCTTGTGAGGCTCTGCTG 1384
Db 1603 TGTACCCGAGCATTAATGATTAAGTACAAAGTGGGCCCGACACTATATAGTGCAGACCCCT 1662
QY 1385 AAAAGTGTGCGAGT---GTGTTCTGGGAGGCCCCCGAGGTCACACAGGTTCCGTA 1441
Db 1663 CGGAGTCCCAAGATACGGCGGCTGTATGTAGCCCCCGAGGTCGGGACCGGTTCCGTC 1722
QY 1442 GGTGTGCGCCCGGCTTCTTCAAGACAGTCTTGTAGAGGCTGTGTAGAGCTTCTGTAAGCGTG 1501
Db 1723 GCTCCCGCCAGGCTTCTTCAAGACCGTGTCTGAGACCTTACTGAAGCTACGCGGAGG 1782
QY 1502 TCGTGTCTGAATGAAGATATCTCTCGGATACCCAGAAATATCGACTGTGTGATGAAA 1561
Db 1783 TAAAGGAGAGATGAAGAGTTCCTCGCTCGACGCCCGAGTACAGGCTCTACGATGAGC 1842
QY 1562 GGCAGAGAGGCTTGAAGCTTCTTGAAGCTTCTTGAAGCTTGTACGCTACATGGGTTGAGAGCGG 1621
Db 1843 GCCAGAGGCGCTCAGAGTCTTTCGCAACCGAGCTATGCTACATGGGGTGGAGCGATG 1902
QY 1622 CTAGGTGTATGTCAGGAGTGGCAAGGCTGTCTACGCGCTTGGGGTAGCACCTCATAC 1681
Db 1903 CCGCTGCTACTGCACACGCTGCGCGAGGCTGTCTACAGCTTGGGCTGTAACCTTATAC 1962
QY 1682 GCACCGCATCAACATAGCTCTGTAACTAGGCTCTCAAGGTGATCTACGGTGAACAGAT 1741
Db 1963 TGACAGCTATCGAGTATGCCAGGAAGCTCGGCCCTAAAGGTTATATATGAGAGACACCGACT 2022
QY 1742 CGCTTCTGCTGACTATGATCCCGAGAGGTCGAAATTTTCATCAAAATATATAAGAGG 1801
Db 2023 CCTCTCTGCTGCTATGACAGAGAGGTTGAGAGGCTGATAGAGTGTGTGCGAGGCT 2082
QY 1802 AGCTGGGTTTCGAAATCAAGCTAGAGAGGTTGACAAAGCTTATTTCTTACAGAGGCTA 1861
Db 2083 AGCTGGGCTTTGAGATAAAGATAGACAGATCTACAGAGAGGTTCTTCTACGAGGCTA 2142
QY 1862 AGAAGAGTACGCTGCGCTTCTCGAGAGCGAGCTATAGATATGTCGTTTTCGAGGCTG 1921
Db 2143 AGAAGCGCTATGTAGGTCTCTCGAGAGCGAGCTATAGACATCGTGGGCTTTGAAGCAG 2202
QY 1922 TACGTGCGGCTGCTGTGAACTCGCAAGGAGGTTGAGAGTGTGCAAGTGTGCAAAATAGTAT 1981
Db 2203 TCCGCGCGGCTGCTGCTGAGGCTGCTGAGAGGTCGAGAGAGGCGGCTGAGATAGTGT 2262
QY 1982 TGAAGACGAGTGAAGTGAACAGGCTGTAGAGTACGTACAGGAAGATTTGAAAGAGTTCG 2041
Db 2263 TGAATACGGGAGCTGGACAGGCTATAGCTACATAGAGGAGGTTAAATAGCAGCTCC 2322
QY 2042 AGAGGCGAAGGTTCCCTATAGAGAGCTTGTAACTTGTGAAGACCCCTTAGTAGCGCTTG 2101
Db 2323 GCGAGGCGAAGGTCGCAATAACAAAGCTTATCATATGGAAGACGCTGAGCAGAGAGTAG 2382
QY 2102 AGAGGTACACACGAGGAGCCACACGCTGTTGCGAGCGAGAGGATGCTGTGAGAGGCT 2161
Db 2383 AGAGGTACGAGCATGACGCGCTCTCATGTGATGGCTGACGCGCTATGAAGAGGAGGCT 2442
```


908 AGATGCCGAGATAAAGATGAAGACGCTTGAAGAGGTAGCGGAGTACCTAGCGGTATGA 967
 965 AGAAGAGTAAACCGCTTATCATCAATGTGGTGGAGATTCAGACATATTGGGACACCGGA 1024
 968 AGAAGAGCGAGCGTGTGATAATAGTGTGGAGGATACCCGAGTACTGGGATGACGAGA 1027
 1025 AGAAGAGACCACTATTACTGCAATACGCGCGGAGCGATGTCCGCGCTACTTACCGCTTAG 1084
 1028 AGAAGAGCGAGCTGTAGAGCGCTACGCGCTCGACGATGTGAGGCTACCTACGCGCTCG 1087
 1085 CCGAGAGATATTTGCGCTTGTCTATCCAGTTGTGTAATACGTAACAGCTCTCCACTAGACC 1144
 1088 CGGAAAGATGCTTACCGTTTCGCCATACAGCTCTCCACTGTTACGCGGTGTCTCTCGACC 1147
 1145 AGGTAGGTGCGATGAGTGTGGCTTTCGACTTGTGATGTTGATGCTGATACGCGCGGTTTA 1204
 1148 AGGTAGGTGCTATGGCGGTAGAGCTTCCGCTAGAGTGTATCTCATGGTGTGCGGCTACG 1207
 1205 AGATGAAGAGCTTTGTCGGAACCGCGTTGAGCGCCGAGAGAGACTTACCGTGGCGCTA 1264
 1208 ATATGAACGAGCTGGTGCAGAACCGGTTGGAGAGGAGGCGGAGAGCTACAAGGCTGCAG 1267
 1265 TAGTCTTGGCGTGTGAGGCGGTGCGAGGATATAGCCGTAATCCGACTTTAGCTCGA 1324
 1268 TAGTGTAAAGCTCTCAAGGGAGTCCATGAGAAATGTTGTGGTCTCGATTTCACTTCCA 1327
 1325 TGTATCCCAACATCATGATAAAGTACAATGTTGGTCTCGACACGCTTGTGAGCGCTGGTG 1384
 1328 TGTATCCCGAGCATATGATGAATGAATACAACTGGGCGCCGACACTATAGTTCGACGACCCCT 1387
 1385 AAGAGTGTGCGAGT---GTGGTGTGCGAGGCGCCGAGGTCAAGCACAGTTCCGCTA 1441
 1388 CGGAGTGCCTCAAGTACGCGCGCTGCTATGTAGCGCCCGAGGTCCGCGTTCGCTC 1447
 1442 GGTGTGCGCGCGCTTCTTCAAGACAGTTCCTTGAAGGCTGTAGAGCTTTCGTAAGCGTG 1501
 1448 GCTCCCGCGCAGGCTTCTTCAAGACCGTCTCGAGAACCTACTGAGCTACGCCACAGG 1507
 1502 TGGGTGCTGAAATGAAGAGATATCTCCGAGTAGCCGAGAACATATCGACTTTGGATGAAA 1561
 1508 TAAAGGAGAGATGAAGAGTTCGCGCTGACAGCGCCGAGTACAGGCTCTACGATGAGC 1567
 1562 GGCAGAGGCGTGTGAAGTTCCTGCAACGCTAGTTTACGCTACATGGTTCGAGCGCG 1621
 1568 GGCAGAGGCGCTCAAGTTCCTTGCAGACCGAGCTATGGCTACATGGGTTGAGCGCATG 1627
 1622 CTAGGTGTATTCAGGAGTGTGCGAAAGGCTGTCAAGGCTTGGGTAGCGCACTCATAC 1681
 1628 CCGCTGTACTGCAAAACGCTGCGCCGAGGCTGTCAAGCTGGGCGCGTAACTTATAC 1687
 1682 GCACCGCATCAACATAGCTCGTAACTAGGCTCAAGGTGATCTACGTTGACAGAT 1741
 1688 TGACAGCTATCGAGTATGCGAGAGCTCGGCTAAAGTTATATATGAGACACCGACT 1747
 1742 CGCTCTCTGCTGACTATGATCCGAGAGGTGGAAAATTTTCATCAAAATATAAAGGAGG 1801
 1748 CCCTCTCTGCTGCTATGACAAAGAGAGTGTGAGAGCTGATAGATTTGTGCGAAGG 1807
 1802 AGCTGGGGTTCGAAATCAAGCTAGAGAGGTGTACAAACGCTTATTTCTTACAGAGGCTA 1861
 1808 AGCTGGGCTTTGAGATAAAGATAGACAAGATCTACAAAGAAAGTGTCTTTCACGAGGCTA 1867
 1862 AGAAGAGTACGCTGCGCTTCTCGAGGACCGAGTATAGATATGTCGGTTTCGAGGCTG 1921
 1868 AGAAGCGCTATGTAGGTCTCTCGAGGACCGAGCTATAGACATCTGTTGGGCTTTGAGCAG 1927
 1922 TACGTGGCGATTTGTTGTAATCTCGCAAGAGGTTTCAGACTAAGGTTTTCGAAATAGTAT 1981
 1928 TCCGCGCGACTGTTGCGAGCTGCTAAGGAGTGCAGGAGAGGCGGCTGAGATAGTCT 1987
 1982 TGAAGACAGTGTAGGTGAACAGGCTGTAGAGTACGTCAGGAAGATTTGTAAGAGTTGG 2041

1988 TGAATACGGGGAACGTGGACAAAGGTATTAAGCTACATAAGGAGGTAAATAAGCAGCTCC 2047
 2042 AGAGGGCAAGGTTCCCATAGAGAAAGCTTGTAAATCTGGAAGACCCCTTAGTAAGGCTCTTG 2101
 2048 GCGAGGGCAAGGTGGCCAAATAACAAAGCTTATCATATGGAAGACGCTGAGCAAGAGTAG 2107
 2102 AGAGTACACACAGGAGGACCAACAGCTGCTGTGAGCGAAAGAGGATGCTGTACAGAGCT 2161
 2108 AGGAGTACGAGCATGACGCGCTCATGTGATGCTGACGCGGTATGAAGGAGGCGAGCT 2167
 2162 ACCGGTACAGCCAGCGACAGATAGGATATGTAATAGTGAAGGCTGTGCGCGTATCA 2221
 2168 ACAGAGTGTCTCCCGCGATAGGTGGCTACGTCAATAGTGAAGGTAGCGGAGTGTGT 2227
 2222 GTCAAGAGCATGCGCCATACCTTCATGGTCAAGGATCTTAGCCAGATAGACGCTGACT 2281
 2228 CCAGCAGGCGCTACCCCTACTTTCATGGT---TGATCCATCGACCATCGACGCTCACT 2284
 2282 ATGTTACCAACCAATCATCCCGCTGCTGATGAGATATCTGGGCTACTTTCGATCAGCG 2341
 2285 ATATTGACCAACCATAGTAGTGGCGCTGCTCTGAGGATACTCTCTTCTTTCGAGTCCCG 2344
 2342 AGAAGAGCTGAAAGCAAGTCAACTGGGCAAGAGACTCTCTTCGACTTCTTAGCAGA 2401
 2345 AGAAGAGCTGAAAGGCGGCTACGGTGCAGAGAGGCTCTTCGACTTCTTCGCTCA 2404
 2402 AGAGCAAG 2409
 2405 AGAATAG 2412

RESULT 10

AAQ73843

ID AAQ73843 standard; DNA; 2430 BP.

XX AAQ73843;

AC AC

25-MAR-2003 (updated)

DT 20-JUL-1995 (first entry)

XX P. abyssi DNA polymerase DNA sequence.

Polymerase chain reaction; PCR; amplify; primer; probe;
 Pyrodictum occultum; DNA polymerase; P. abyssi; condensation;
 DNA sequencing; nick translation; reverse transcription;
 heat stability; heat resistant polymerase; DNA helix stabiliser; ss.

XX Pyrodictum abyssi.

OS Key Location/Qualifiers

XX CDS 1..2412

XX /*tag= a

XX /product= DNA polymerase

XX EP624641-A2.

XX 17-NOV-1994.

XX 02-MAY-1994; 94EP-0106811.

XX 14-MAY-1993; 93US-0062368.

XX (HOFF) HOFFMANN LA ROCHE & CO AG F.

XX Gelfand DH, Wang AM;

XX WPI; 1994-350781/44.

XX P-PSDB; AAR60882.

New heat stable nucleic acid polymerase from Pyrodictum species
 - and relates DNA, vectors and transformed hosts, resistant to
 denaturation at 100 deg.C, useful for nucleic acid amplification,
 in DNA sequencing etc.

607	GTAAATAGTGTGATAGCTGTGAGACATGACGATGGCGATGAGAGTGTCTATPTCATTTGCAGAGGCG	666
2Y		
691	GTGGTTATGTATTAGCCATAAAGACAAACGATGGGCACGAGAGAGTGTTTTGAAGCTAGTGGG	750
Db		
667	AAAGACGATCGAAACCCGATACGCGAGTGTCTGAGTACGCTGAAGAGGTATGACCCCGAC	726
2Y		
751	AAAGACGATAGGGGGTGTGCGCGCTTTTGTGGATTTTATCAGAGTTTATGACCCCGAT	810
Db		
727	ATATATGTGGGTATATAACAAATCATTTTCAATTTGGCTTATCTTTTGAAGGCGCGCCGC	786
2Y		
811	GTAAATTTGGCTTATAATTCCAAAGCGCTTTGATTTGGCTTTATTTAGTAGAGCGTGGAAA	870
Db		
787	ATCCTTAGGCATAAAGCTTGTATGTGACTAGAAGAGTTGGCGCGAGGCCACCATAGCGTA	846
2Y		
871	GCGCTCGGCGTCCGCTCAAGGT---TGATAGACTGAGCAATCCCCCTCAACAGAGCGTG	927
Db		
847	CATGGGCAAGCTCTCTGTCCCTGGCAGGCTTAACGTAGATCTGTACAGTATGTCGAGAG	906
2Y		
928	TATGGACACTGGTCTATCGTTGGTAGACAAATGTAGTCTCTATACACTCGTGGAAAG	987
Db		
907	ATGCCAGAGATCAAGATAAAGAGTCTCGAGGAGTCTCGAGAGTCTAGCGGTGATGAAG	966
2Y		
988	TTCCCGAGATTAAAGCTAAAGACTCTTCACCGGTGTCTGAATATTTTGTGTATATGAAA	1047
Db		
967	AAGNCTGAACGCGTTATCATCATTAATTTGGTGGAGATCCAGACTATTTGGGACGACCGAG	1026
2Y		
1048	AGAGAAGAGAGGGTGTGTGATCCGAGCCACAAGATTTATGAATACTGGAAAGACCCCAAT	1107
Db		
1027	AAAGACACCTATTACTTGCATATACGCGCGCAGCATGTCCGCGCTACTTTACGCTTAGCC	1086
2Y		
1108	AAAAGACCTTACTTTAAGCGGTATGTTCTAGACGACGTACGATCCACTCTAGGACTTGGC	1167
Db		
1087	GAGAGATATTGCGGTTTCTATCCAGTTTCTGATACGTAAACAGTCTCCCACTTAGACCG	1146
2Y		
1168	GACAGCTCTACCGTTTTTAATACACTGTCTCTGTATCTGGGCTACCGCTTGGATCAG	1227
Db		
1147	GTAGGTCCGATGAGTGTGGCTTTTCGACTTGAATGGTACCTGATPACGCGGGGTTTAAG	1206
2Y		
1228	GTGGACGCGCGAGCGTGGGCCAACAGGCTAGAGTGGATGCTTCTTAGGTACGCGTACC	1287
Db		
1207	ATGAAGAGCTCTGTGCCGAACCGTTGAGCGCCACAGAAGACCTTACCGTGGCGCTATA	1266
2Y		
1288	CTGGCGAGGTAGCTCCGAACAGAGAGAGGAGGTACGAAACCTATTAAGGGGGCCATC	1347
Db		
1267	GTCTTTGAGCGGTTGAGAGCGGTGCACGAGAATATAGCGGTACTCTGACTTTTAGCTCGATG	1326
2Y		
1348	GTGCTTGAGCAAAAGCCCGGATGTACGAAGACGTGCTTGTACTTGTACTTCTCTTCCATG	1407
Db		
1327	TACCCAAACATCATGATAAAGTACAATGTGGTCTGACACGCTTGTGAGGCGCTGCTGAA	1386
2Y		
1408	TACCTTAACATCATGTATGATGAGTACAACTATCTCGCGGATACGTACTTAGAGCCCGCGAG	1467
Db		
1387	AAGTGTGGGAGTGTGGTTCTGGGAGGCGCCGGAGGTCAAGCACAGTTCCGTAGTGT	1446
2Y		
1468	CCGATCCCGCCGAGGGTGTAAACGTGCGCCCGGAGTGGGCGCATATGTTTAGTGAAGC	1527
Db		
1447	CCGCCCCGCTTCTTCAAGACAGTCTCTGAGAGGCTGTTTAGAGCTTCGTAAAGCTGTGCGT	1506
2Y		
1528	CCACCAAGCTTCGTCCCTCAGGTGTGAAGAGCTTGGTGGAGCTTAGAAGCGGTAAAGA	1587
Db		
1507	GCTGAATGAAAGATATCCTTCGGATAGCCAGCAATATCACTGTGTGGATGAAGGCAG	1566
2Y		
1588	GAGAGGCGAAGAAATACCCCCAGATTCCGCCAGAGTTTAAGATCTTTAGACGAGACAA	1647
Db		
1567	AAGCGGTTGAAGTTCCTTGCAACCGTAGTTACGGCTACATGGGTTGGAGCGCGTAGG	1626
2Y		
1648	CGTCCCTCAAGGTTATGGCCATGCCATTTATGGATACTTTGGGCTGGTGGGCGCCCG	1707
Db		
1627	TGCTATTGGCAGGAGTGGCCAAAGGCTGTCAAGGCTTGGGTAGGACCTCATACGCA	1686
2Y		
1708	TGGTATAAGCGGGAGGTAGCCGAGTCTGTGACGGCTTTTGGCAGAGCGATTTCTAAAGAC	1767
Db		
1687	GCCATCAACATAGCTCGTAACTTAGCGCTTCAAGGTGATCTACGGGTGACACAGATTCGCTC	1746
2Y		

1768	DB	GTATTGAACAAGCTAGAAAGGTGGGCANTGTGGTTGTATATGGCATTACAGACACCTTA	1827
1747	QY	TTCGTGACCTATGATCCGGAGAGAGGTGGAAATTTTCATCAAAATTTATAAGGAGGAGCTG	1806
1828	DB	TTTGT---CAAAAAACATGGAGACCTGGACAACTGATCAAGTACGTGGAGGAGAAGTAC	1884
1807	QY	GGCTTCGAATCAACTACAGAGAGGTGTACAAACGCTTATCTTTACAGAGGCTAAGAG	1866
1885	DB	GGCATAGACATTAAGGTGACAGGATTACGCCAAGGTGCTTTTACGGAGGCTAAGAG	1944
1867	QY	AGGTACGCTGGCTTCTCGAGACGACGCTATAGATATTCTCGTTTCGAGGCTGTACGT	1926
1945	DB	AGGTACGCTGGCTTATTGAGAGATGGGCGTATAGATATGTGTGGGATTCGAGGTCGTGAG	2004
1927	QY	GGCGATTGTGTGACATCTGSCCAAGGAGGTTCAGAGCTAAGGTGTGCGAATAGTATTTGAAG	1986
2005	DB	GGAGCTGGNGTGAACCTTCCCAAGACGTTCCGCTAAGAGTTATAGAGATTATCTTAAG	2064
1987	QY	ACGGGTGAGGTGAACAAGGC-----TGTAGAGTACGTCAGGAAGATTGTGAAA	2034
2065	DB	TCAAGAGATATTGTGGAGCTAGACATGGCGTTATAAAATATATTAGAGAAATAATAGAA	2124
2035	QY	GAGTTTGGAGGCGGCAAGGTTCCCATAGAGAAGCTTGTAATCTCGAAGACCCCTTAGTAAG	2094
2125	DB	CGACTGAAAAATTACAGTTTAACTTGATGNTTAAATATATGGAATCTCTAGACAAA	2184
2095	QY	CGTTCTTGAGAGTACAAACGGAGGCAACACAGTCGTTGTCAGGAAGAGGATGCTGTCA	2154
2185	DB	GAGTTAGACGAATATAAGCGGTATCCGCTCATGTTCACGTCGCGAAATTCCTCAAGAGA	2244
2155	QY	GCAGGCTACCGGTAAAGCCAGCGCAAGATAGGGTATGTAATAGTGAAGGTCGTGCG	2214
2245	DB	CACGGCTATCGATGGGCAAGGCATCAGATAGCTATGTCTATGTTAAAGGAGGCGAA	2304
2215	QY	CGTATCAGTCCAAAGACATGGCCCATCTTCATGTGTCAGGATCTCTAGCCAGATAGACGTG	2274
2305	DB	AAAGTATCTGAGAGGGCGTTGCGGTATATATCTCTTGACGATATAAAAAAGATAGATATC	2364
2275	QY	ACCTTACTATGTTGACCAACCAATCATCCCGGCTGCATTTGAGAACTACTGGGCTACTTTGGC	2334
2365	DB	GACTACTACATAGAGAGACAGATAATACACGGCGTTTGAGATAGCAGAGGTTATAGGG	2424
2335	QY	ATCACCGAGAAAGCTGTGAACCAAGTCGACTCGCCAGAGAGCTCTCTTCGACTTTCTTA	2394
2425	DB	GTGAAGAGGTCGGATCTTAAAAACGGGGAAA---TGGAAAGGTCGCTCTTGATTTTCTA	2481
2395	QY	GCCAAGAGAGCAAGT	2410
2482	DB	AGTTAGAAAACTACGT	2497

RESULT 12
AAAX27284
ID AAAX27284 standard; DNA; 2367 BP.

AA
AC
AAX27284;

DT 02-JUN-1999 (first entry)

DE A. lithotrophicus DNA polymerase 5PY1 coding sequence.

XX DNA polymerase; thermophilic bacteria; DNA synthesis; ss.
KW

XX Archaeoglobus lithotrophicus. OS

XX PN WO9907837-A1.

XX PD 18-FEB-1999.

XX PF 06-AUG-1998: 98WO-US17152.

XX 06-AUG-1997: 97US-0907166.

1687 GCCATCAACATAGCTCGTAACTAGGCCCTCAAGTGATCTACGGTGACACAGATTGGCTC 1746

XX (DIVE-) DIVERSA INC.
 XX Callen W, Mathur B;
 XX WPI: 1999-180490/15.
 DR P-PSDB; AAY00937.
 XX
 XX DNA polymerases from extremely thermophilic bacteria - useful for
 XX DNA synthesis
 XX
 XX Claim 3; Fig 3; 72pp; English.
 XX
 XX This sequence encodes a DNA polymerase of the invention, that was
 XX isolated from a thermophilic bacteria. The polymerases are used in DNA
 XX synthesis and as immunogens to raise antibodies (useful for affinity
 XX purification and to screen for related enzymes). Fragments of the DNA
 XX encoding the polymerases are used as probes to isolate related or
 XX full-length sequences and to produce the recombinant polymerases. The
 XX polymerases catalyze DNA synthesis by the addition of deoxynucleotides to
 XX the 3' end of a polynucleotide chain, using a complementary
 XX polynucleotide strand as a template. The polymerases have optimum
 XX activity at over 60 deg. C and can renature and regain activity after
 XX exposure to temperatures above 70 deg. C.
 XX
 XX Sequence 2367 BP; 799 A; 417 C; 612 G; 539 T; 0 other;

Query Match 11.6%; Score 279; DB 20; Length 2367;
 Best Local Similarity 50.9%; Pred. No. 1.9e-75;
 Matches 889; Conservative 0; Mismatches 795; Indels 63; Gaps 7;
 QY 530 TTCGCGATCTAGGATCTACGCTCGCTTCGATTTGATGTTTATACGACGAGGTCGCCG 589
 DB 488 TTCCCGAACTCAGGTTATGCGTTGTTGTCGGAATGCTCTCAGAGGTTGTTATGCCG 547
 QY 590 GTCCAGAGCGCGATCTCTGTTATAGTGTAGTGTGAGATCTGACGATGCGATGAGGTC 649
 DB 548 ATCCAGAGAAAGATCTATCATGATCATCAATTAATTAATCGGCTGAATACGAGGAATCC 607
 QY 550 TATTCATTTGAGAGGCGAAGACCATCGAAACCGATACCGAGTTGTAGATGACGTA 709
 DB 608 TCAACGGG-----ATAACGAGAGAGATGCTTACAGATTTGCAAGTAATTC 658
 QY 710 AGAGTATGACCCGACATATAGTTCGTTTATACAAATCAATTCGATTCGCTTATC 769
 DB 659 GCGATATTGATCCGACATTAATGTTGATACAAATCAGGACAGCTTGACTGCCCTATA 718
 QY 770 TTTTACGCGCGCGGATCTCTAGCATTAAGCTTGATGTCACTAGAGATTGGCGCG 829
 DB 719 TCAAGAGAGAGCTGAGAACTGAGGGTTAAGCTTGACATCGGAAGAGATAGAGCGAAC 778
 QY 830 AGCCCAACCATAGGTATGTCGTCGTCGCTCTCTGCTGCGAGGCTTAACTGATCTGT 889
 DB 779 TGGCTATCAGGGGAGAGACCAA-----GATTGCTGCGAGGTTGAACGTGATCTCT 832
 QY 890 AGCACTATGCCGAGAGATGTCAGAGATCAAGATTAAGATCTCAGGAGGTCGACAGT 949
 DB 833 ATGATATTGCAATAGGAGTCTCGATGTAAGAGTAAGAGAGCTGCAAAAGCTGTCAGAGT 892
 QY 950 ATCTAGGCGGTGATGAAGAGAGTGAACGCTTATCAATTTGTTGGAGATTCAGACT 1009
 DB 893 TTCTGGTAAAGAAATAGGCTTGCAGATATTGAAGCGAGGATATCTCAACACTGGA 952
 QY 1010 ATTGGGACGACCCGAGAGACCACTATTACTGCAATACGCGCGGAGCATGTCGCGG 1069
 DB 953 CATCGGGGACAGGGAACG-----TAATCAATATCTCCCGCAGGACATCTCTC 1003
 QY 1070 CTACTTACGGCTTACGCGAGAGATATTGCGGTTTGTCTATCCAGTTCTCTAGCTAACAG 1129
 DB 1004 ACAGCTATTCTACTGAGAAATGCTGCCCAATGCAATAGCACTTCCAGATGATAC 1063
 QY 1130 GTCTCCCACTAGACGAGTGTGCGATGAGTGTGGCTTTGCACTTGAATGTACCTGA 1189

DB 1064 GCATACCTCTCGATGATGTGACACAGGAGGAGAGGTAAGCAGGTTGAGTGGCTGCTGT 1123
 QY 1190 TACGCGCGGCTTTAGATGAAGAGCTTGTGCGAACCGGTTGAGCGCCCAAGAGA 1249
 DB 1124 TAAGCGAAGCACAACAACTTGGCAACTTGCACCCCAACCCAGAGAG---ATGCCGACA 1180
 QY 1250 CTTACCGTGGCGCTATAGTTCTTGTAGCGGTTGAGAGGCGTGCACGAGAAATATATCCGTAC 1309
 DB 1181 GCTATGAAGGAGCATTCGTGCTCGAGCCGCGCAAGAGGATTGCATGAGAACGTAATCTGCC 1240
 QY 1310 TCGACTTGTAGTGTACCTACCAACATCATGATTAAGTACAAATGTTGGTCTCTACACGC 1369
 DB 1241 TGGACTTTCGTCCTATGATCCCTCAATATGATTTTCATACACATCAGCCCCGACACGC 1300
 QY 1370 TTGTGAGGCTGTGTAAGAGTGTGCGAGTGTGTTGCTGGAGAGCCCGGAGGCTCAAGC 1429
 DB 1301 TTCT-----AATAGCAAAATGCGACGATTGCAATGTAGCGCGGAGGTGGGC 1348
 QY 1430 ACAGGTTCCGTAGGTTCCGCGCGCTTCTTCAACACAGTCTTGTAGAGGCTGTAGAGC 1489
 DB 1349 ACAATTCAGGAACATCTCTGATGTTTTCCTCAAAAGAAATCTCAAAATGCTGATTGAGA 1408
 QY 1490 TTGCTAAGGCTGTGCTGCTGAATGCAAGAGTATCTCCGATAGAGCCCAAGATATCCAC 1549
 DB 1409 AAAGAGAGAAATAAAGAGGTTATGNAACACTTGACTACAACTCGCCAGATACAGC 1468
 QY 1550 TGTGTGATGAAGCGAGAGGCTTGAAGGTTCTTGCAAAACGCTAGTTACGGCTACATGG 1609
 DB 1469 TGTCTGATTAAGCAGGCAACGCTGAAGTTCTTACAAACTCGCTTTTACGTTTATCTGT 1528
 QY 1610 GTTGGAGCGCGCTAGTGTGTTTTCAGGAGTGTGCGCAAGGCTCTACGCTTGGGGTA 1669
 DB 1529 GGTGGAGTCTTCGAGATGTTCTCAAGAGTGTGCTGAACTACAACTCAAGCGCATGGGCA 1588
 QY 1670 GGCACCTCATACGACCGCATCAACATGCTGTAACCTAGGCTCAAGGTGATCTACG 1729
 DB 1589 GACACTTTATCAAAACATCTGCAAGAAATTCGAAAGAGCTTGATTTGAAGTGTATATG 1648
 QY 1730 GTGACACAGATTGCTCTTCGTGACCTATGATCCGAGAGGAGTGGAATAATTCATCAAAA 1789
 DB 1649 GGNATACAGATACATCTTTGTTAAAGAGTGGATTGAGCTTGGAGAGCTCAAAAAG 1708
 QY 1790 TTATAAA-----GGAGAGCTGGGGTTCGAAATTCAGCTAGAGA 1828
 DB 1709 AAGTTAAAAAGCTCATAGGTAAACTTTTCGGAAGAGATGCCATACAAATAGAGATGATG 1768
 QY 1829 AGGTGTACAAACGCTTATCTTTACAGAGGCTTAAGAGAGGTACGCTGCTCTCTCGAGG 1888
 DB 1769 AATACTACGAGCAATATCTTCGTTGAA--AAGAAAGGTATCTCTGATTTGACACAG 1825
 QY 1889 ACGGAGTATAGATATTGTGGTTCGAGGCTGTACGTTGCGAATTTGGTGTGAACCTGCCA 1948
 DB 1826 ATGGAAGATAATTTGAAAGGGCTTTGAAAGTCAGAGAGGCGACTGTCGAGCTTGCA 1885
 QY 1949 AGGAGGTTACAGCTAAGGTTTGTGCAATAGTATTGAGACGAGTACAGTGAACAGGCTG 2008
 DB 1886 AGAAGATACAGAGGTTAATAGAAATCTTCTGAGGAAGAAATCTCTGAAAAGCTG 1945
 QY 2009 TAGAGTACGTCAGGAAGTTGTGAAAGTTTGGAGAGGCGGAGGTTCCCATAGAGAGC 2068
 DB 1946 CTGAGTATGTGAAAGGAGTCATAGAGGAGATAAAGCGAGGCAAAAATTCGCTTGAAGATT 2005
 QY 2069 TTGTAATCTGGAAGACCTTTAGTAAAGCTTTGAGGAGTACACACGAGGACCAACAGC 2128
 DB 2066 TAAAGCTGCCATGAAGGCGCAAGAGAGGAATAGTATACACAATCGCTCAAGGTTG 2125
 QY 2189 GGTATGTAATAGTGAAGGTTGGCGGTTATCAGTCAAGAGCATGCGCATCTCTATGG 2248
 DB 2126 GTTTTGTGTTTCAAAAGGTTGGGGAACATAGGTGATAGGGCTTTTCCATCTGATCTGA 2185

QY 1790 TTATAAA-----CGAGGAGCTGGGTTTCGAAATCAAGCTAGAGA 1828
 Db |||
 QY 1709 AAGTTAAAGAGCTCATAGGTAAACTTTCCGAAGAGATGCCAATACAAATAGAGATAGATG 1768
 Db |||
 QY 1829 AGGTGTAACAAGCTTATTTTACAGAGCTAAGAGAGTAGCTGGCTTCTCGAGG 1888
 Db |||
 QY 1769 AATACTACGAGACAATATTTCTCGTTGAA---AAGAAAGGTATGCTGGATTGCACACAGG 1825
 QY 1889 ACGAGCTATAGATATTTGTCGGTTTCGAGCTGTACGTGCGGATTTGGTGTGAATCTGCCA 1948
 Db |||
 QY 1826 ATGAAGATAAATTTGTAAGGGTCTTGAAGTCAAGAGAGCGGACTGTGCGAGCTTGCAA 1885
 QY 1949 AGGAGGTTTCAGACTAAGGTTGTGCAATAGTATTGAAGACGAGTGAAGTGAACAAGGCTG 2008
 Db |||
 QY 1886 AGAAGATACAGAAAGGTGTAATAGATATCTTGAAGGAAAGATCTGAAAGAGCTG 1945
 QY 2009 TAGAGTACGTCAGGAAGATTGTGAAGAGTTGGAGAGGCGCAAGGTTCCCATAGAGAAGC 2068
 Db |||
 QY 1946 CTGAGTATGTGAAGGAGTCAATAGAGGAGATAAAGCGCAGCAAAATTCGCTTGAAGATT 2005
 QY 2069 TTGTAATCTGGAAGACCTTAGTAAAGCTTCTGAGGAGTACACAAAGGAGGCCACACAG 2128
 Db |||
 QY 2006 ATATCATCTCAAGGGATTACAGAGAAACCATCAAGTACGAGAGTATGCAAGGCTCAGC 2065
 QY 2129 TCGTTGAGCGAAGAGGATCTGTCAAGAGGCTACCGGGTAAGCCGAGCGCAAGATAG 2188
 Db |||
 QY 2066 TAAAGCTGCCATGAAGCGGCAAGAGAGGAAATGATATACAACTCGCTCAAGGTTG 2125
 QY 2189 GGTATGTAATAGTGAAGGTTGGCGGTATCATAGTCAAGAGATGCCATCTTCATCG 2248
 Db |||
 QY 2126 GTTTTCTGCTTCAAAAAGGTTGGGAAACATAGGTGATAGGGCTTTTCCATCTGATCTCA 2185
 QY 2249 TCAAGGA 2255
 Db |||
 QY 2186 TAGAGGA 2192

RESULT 14

ABX14887
 ID ABX14887 standard; DNA; 2367 BP.

AC ABX14887;

DT 08-APR-2003 (first entry)

DE DNA encoding Archaeoglobus thermostable DNA polymerase protein.

KW Gene; ds; thermostable; DNA polymerase; DNA repair;

KW polymorphism identification.

OS Archaeoglobus lithotrophicus.

FH Key Location/Qualifiers

FT CDS 1..2367

FT /*tag= a

FT /product= "A. lithotrophicus DNA polymerase"

XX US2002132243-A1.

PN 19-SEP-2002.

PD 06-SEP-2001; 2001US-0948369.

PF 06-AUG-1997; 97US-0907166.

PR 07-SEP-1999; 99US-0391340.

PR 06-SEP-2000; 2000US-0656309.

XX (CALL/) CALLEN W.

PA (MATH/) MATHUR E J.

PA (SHORT/) SHORT J.

PI Callen W, Mathur EJ, Short J;

XX

DR WPI; 2003-182285/18.

DR P-PSDB; ABG73156.

XX

PT Novel DNA polymerases having increased activity and stability at increased pH and temperature, useful for DNA sequencing, amplification and incorporating non-natural nucleotides or nucleotide analogs -

PS Claim 1; Page 38-39; 81pp; English.

XX

CC This invention relates to a novel purified polypeptide derived from Pyrobolus fumari which has thermostable DNA polymerase activity. The protein of the invention is useful for catalysing the formation or repair of a nucleic acid sequence, for comparing a sequence to a reference sequence and for identifying polymorphisms. The nucleic acid sequence of the invention is useful for preparing cDNA from mRNA, and in an amplification process of a double-stranded DNA molecule. The nucleotide sequence is also useful for incorporating non-natural nucleotides or its analogues into a DNA molecule, by contacting a polypeptide encoded by the nucleic acid with a DNA template in a PCR amplification reaction. The nucleotides which may be used for this are inosine, 2-aminopurine or 5-methylcytosine. A nucleic acid probe derived from the nucleic acid sequence of the invention is useful in chromosome walking procedures to identify clones containing genomic sequences located adjacent to a sequence of the gene encoding the DNA polymerases of the invention. Such methods allow the isolation of genes which encode additional proteins from the host organisms. The polymerases of the invention have increased activity and stability at increased pH and temperature, and high processivity. The present invention represents a DNA sequence encoding an Archaeoglobus lithotrophicus thermostable DNA polymerase protein of the invention.

XX Sequence 2367 BP; 799 A; 417 C; 612 G; 539 T; 0 other;

Query Match

Best Local Similarity 11.6%; Score 279; DB 25; Length 2367;

Matches 889; Conservative 0; Mismatches 795; Indels 63; Gaps 7;

QY

530 TTCGGATCTTAGGATACCTCGGTCGATATTGAAGTTTATGACGAAGCGGTCGCCG 589

Db

488 TTCGGAACTCAAGGTTATGCGCTTGTATTGCGAAATGCTCTCAGAGTTGGTATGCCG 547

QY

590 GTCCAGAGCGGATCCTGTATAGTAGTCTGTGAGACTGAGATGGCATGAGTGC 649

Db

548 ATCCAGAGAAAGATCCTATCATGTCTATTAATTCGGGTGAATACAGGAAATCC 607

QY

650 TATTCATTTGACAGGGGCAAGACGATCGAAACCGATACCGAGTTTGTAGAGTACGTGA 709

Db

608 TCAACGGTG-----ATAACGAGAGAGATTGCTTACCAGATTGTCTAGATATTC 658

QY

710 AGAGTATGACCCGACATATATGTTCGTTATATCAACAATCATTTTCGATTGGCTTATC 769

Db

659 GCGATATTGATCCCGACATATATAGTTGGATACATCAGGACAGCTTTGACTGGCCCTATA 718

QY

770 TTTTGGCGCGCCCGCATCCTAGGCATAAAGCTTGATGTGACTAGAGAGTTGGCCCG 829

Db

719 TCAGAGAGAGCTGAGAACTGAGGTTAGCTTGACATCGGAAGAGATAGAGCGGAAC 778

QY

830 AGCCACCACTAGCGTACATGGGCACGCTCTCTGTCCCTGGCAGGCTTAAACGTAGATCTGT 889

Db

779 TGGCTATCAGGGGAGGAAGACCAA-----GATTGCTGGCAGGTTGACGTGATCTCT 832

QY

890 ACGACTATGCCGAAGAGATCCAGAGATCAAGATAAAGATCTTCGAGAGGTCGCGAGT 949

Db

833 ATGATATTGCAATGAGGAGTCTCGATGTAAAGGTGGAAGAGCTCGAAACGTTTCGAGT 892

QY

950 ATCTAGCGTGTATGAAGAAGAGTGAACCGCTTATCATCAATTGGTGGGAGATTCAGACT 1009

Db

893 TTCTGGGTAGAAATATAGAGCTTGCAGATATTGAAGCGAAGGATATCTACAGCCTGGA 952

QY

1010 ATTGGGACGACCCGGAAGAGAGACCACTATTACTGCAATACGCGCGGAGATGTCGCG 1069

Db

953 CATCGGCGACAGGGAAGCG-----TAATCAATTAATCTCCGCGCAGGACATCTGCG 1003

1070 CTACTACGGCTTACCCGAGAGATATTGCGGTTGCTATCCAGTTGTCGTACGATACAG 1129
Db ACACGTACTTATAGCTGAAGAATTCGTCGCAATGCAATTCAGAACTTCCAGAAATGATAC 1063
QY 1130 GTCCTCCACTACGAGGTAGTGGATGATGTTGGCTTTCGACTTGTATGTTGCTGCT 1189
Db 1064 GATACCTCTCGTATGATGACAGGAGCGGAGGTTAAGCAGGTTGATGTTGCTGCTGCT 1123
QY 1190 TACGCGCGCGCTTAAAGATGAAGAGCTTGTGCGAACCCTGTTGAGCGCCCGAGAGAGA 1249
Db 1124 TAAGCGAGACACAACTTGGCGAATTCGCAACCCAGAGAG--ATGCGCGACA 1180
QY 1250 CTTACCGTGGCGCTATGTTCTTGAGCGGTTGAGAGCGCTGCAAGAGATATAGCGTAC 1309
Db 1181 GCTATGAAGAGCATTCGCTGCGAGCCCGAAGAGGATTGCAATGAGAACTGATTCGCC 1240
QY 1310 TCGACTTTAGCTTCGATACCCCAACATCATGATAAAGTCAAAATGTTGGTCTCTGACACGC 1369
Db 1241 TGGACTTTGCGTCCATGATCCCTCAATATGATTTCTATCAACATCAGCCCGACACGC 1300
QY 1370 TTGTGAGCGCTGTAAGAGTGGCGAGTGGTGTGCTGGAGCGCCCGAGGTCAGC 1429
Db 1301 TTGT-----ATAGCAAAATCGACGATTCGAATGTAGCGCCCGAGGTTGGGC 1348
QY 1430 ACAGGTTCCGTAGGTGTCGCCCGGCTTCTTCAAGACAGTTCTTGAAGAGCTGTTAGAGC 1489
Db 1349 ACMAATTCAGGAACATCTGATGTTTTCBAAGAGATCTCAAAATGCTGATGAGA 1408
QY 1490 TTCTGAAGCGTGTGCTGTAATGAAGAGTATCCCTCGGATAGCCCGCAATATGAC 1549
Db 1409 AAAGAAGAGAAATAAAGAGGTATGAAGACACTTGAATCAACTCCGCAATATACAGC 1468
QY 1550 TGTGATGTAAGAGGAGAGCGTGAAGGTTCTTGAAGAGCTTCTGCAACGCTAGTTAGCGTACATGG 1609
Db 1469 TGCTCGATATAAGCGCAAGCTGAAGTCTTCAAACTCTTGAAGGTTTGAAGTGTATGCTG 1528
QY 1610 GTTGAGCGCGCTAGGTGTTATTCAGGAGTGGCGAAAGGCTGTCAACGCTTGGGGTA 1669
Db 1529 GGTGAGTCTTGGAGATGTTACTGCAAGAGAGTGGCTGAAGCTTACAAACGCGATGGGCA 1588
QY 1670 GGCACCTATACGCCCGCATACATAGTCTGTAACCTAGGCTCAAGGTGATCTACG 1729
Db 1589 GACCTTTATCAAAACATCTGCAAGATTTGCAAGAGCTTGGATTTGAAGTGTATATG 1648
QY 1730 GTGACACAGATTCGCTTCTGACCTATGATCCGAGAGGTGCAAAATTTTCATCAAAA 1789
Db 1649 GCGATACAGATAGCATCTTGTATAAAGATGATTCAGCTTGAAGAGCTCAAAAG 1708
QY 1790 TTATAA-----GGAGGCTGGGTTGCAATCAGCTAGAGA 1828
Db 1709 AAGTTAAAGCTCATAGTAACTTTTCGGAAGAGATGCCAATCAAAATAGATAGATG 1768
QY 1829 AGGTGTACAAAGCTTATCTTTACAGAGGCTTAAGAGAGTACGCTGGCTTCTCGAGG 1888
Db 1769 AATACTACGAGCAATATCTTCGTTGAA--AAGAAAGTATGCTGATTCAGACAGG 1825
QY 1889 ACGAGCTATAGATATTCGTTTCGAGGCTGTACGTGGCGATTTGGTGAATCCGCCA 1948
Db 1826 ATGGAGAGATAATTGTAAGGGTCTTGAAGTCAAGAGAGGCGACTGTGTCGAGCTGCA 1885
QY 1949 AGGAGGTTTCAGACTAAGGTTGCGAAATAGTATTCAGACAGCTGAGGTGAACAGCTG 2008
Db 1886 AGNAGATACAGAGGTGTAATAGAAATCTTCTGAGAGAAAGAAATCTGAAAGAGCTG 1945
QY 2009 TAGAGTACCTCAGAGATTTGTAAGAGATTTGAGAGGCGCAAGGTTCCCATAGAGAGC 2068
Db 1946 CTGAGTATGTGAAGAGGATCATAGAGAGATTAAGGCGAGGCAAAATTCGGTTCAGAT 2005
QY 2069 TTGTATCTCGAGACCTTGTAGCGTCTTGAAGAGTATACACAGGAGCCACACG 2128
Db 2006 ATATCATCTACAGGGATTTGACGAGAAACCATCAATCAAGATACAGAGTATGACGCTCAGC 2065

QY 2129 TCGTTCAGCGAAGAGGATGCTGTACAGAGCTTACCGGTAAGCCCGGCAAGATAG 2188
Db 2066 TAAAGCTGCGCATGAAGCGCGCAAGAGAGGAATAGTATACAAATCGGCTCAAAGGTTG 2125
QY 2189 GGTATGTAATAGTCAAGCGTGGTGTATCAGTCAAGAGCATGCCATCTTCATCG 2248
Db 2126 GTTTGTGTTACAAAGGTTGGGACATAGGTAGGGGTTTTTCCATCTGATCTGA 2185
QY 2249 TCAAGGA 2255
Db 2186 TAGAGGA 2192

RESULT 15
AAT88374
ID AAT88374 standard; cDNA; 2376 BP.
XX AAT88374;
XX AC
XX 28-JAN-1998 (first entry)
XX Thermococcus peptonophilus SM-2 thermostable DNA polymerase cDNA.
XX JCM 9654; thermostable; DNA polymerase;
KW nucleic acid sequence amplification; PCR; ds.
XX Thermococcus peptonophilus.
XX OS
XX PH Key Location/Qualifiers
XX FT 1..2325
XX FT /*tag= a
XX FT /product= DNA polymerase
XX JP09252776-A.
XX PN
XX 30-SEP-1997.
XX PD
XX PF 19-MAR-1996; 96JP-0063112.
XX PR 19-MAR-1996; 96JP-0063112.
XX XX (TOYM) TOYOBO KK.
XX WPI; 1997-530149/49.
XX P-PSDB; AAW3112.
XX Thermococcus peptonophilus thermostable DNA polymerase - useful for
XX nucleic acid sequence amplification, e.g. polymerase chain reaction
XX Claim 13; Pages 19-22; 27pp; Japanese.
XX PS
XX CC The present sequence encodes a Thermococcus peptonophilus SM-2
XX CC (JCM 9654) derived thermostable DNA polymerase, which can be used
XX CC for nucleic acid sequence amplification, e.g. PCR.
XX SQ Sequence 2376 BP; 679 A; 546 C; 693 G; 458 T; 0 other;

Query Match 8.5%; Score 204.2; DB 18; Length 2376;
Best Local Similarity 48.5%; Pred. No. 3.5e-52;
Matches 862; Conservative 0; Mismatches 848; Indels 69; Gaps 8;
QY 622 GTGAAGACTGACGATGCGGATGAGTGTATTCATTTCAGAGGCGCAAGACGATCGAAA 681
Db 508 GTGTAATGCGAGACATAGACCTCCCTACCTTACGCTGCTTCGAGAGAGGAG 567
QY 682 CCGATACCGAGTTTGTAGAGTACGTGAAGAGGTATGACCCCGCATATAATAGTCGTTAT 741
Db 568 ATGATAAAGCGCTTCTCAGGATTTGTGAAGAGAAAGACCGGACGTAATAAGCTAC 627
QY 742 AACACATCATTTGATTTGCTGCTTATTTTGGGCGCGCCCGCATCTTAGGCATTAAG 801
Db 628 AACGTTGCAACTTTGACTTTTGGCTTACCTCAAGAGCGCTGTGAAAAGCTTGAATAAAC 687

QY 802 CTTGATGTGACTAGAGAGTTGGCCGAGCCACCACTAGCGTACATGGCGACGTCCT 861
Db |||||
QY 688 TTTCCTCGAAGGACGAGCCCTAAGTTTCAGAGGATGGCCATAGGTTTGT 747
Db |||||
QY 862 GTCCCTGCGAGGCTTAACGTAGATCTGTAGACTATGCCGAAGAGATGCCAGAGATCAAG 921
Db |||||
QY 748 GTCGAGGTTAAGGCGAGGATACACTTTTGTATCTACCCCGTGATAAGACGGACGATAAAC 807
QY 922 ATAAAGAGTCTCGAGGAGTGCAGAGATCTAGGCGTGATGAAGAGAGTGAACCGCTT 981
Db |||||
QY 808 CTCCCAACATATACGCTTGAGGCTGTCTATGAAGCATCTTTGGAAG---CCAAAGGAG 864
QY 982 ATCATCAATGTGGAGATTCAGACTATTTGGACAGCCCGAAGAGAGACCACTATTA 1041
Db |||||
QY 865 AAAGTCTACGCCGAGGATAGCCACCGCTTGGAGAGCGGGGCTTGAGAGGGA 924
QY 1042 CTCATATACCGCGGACAGATGTCCGCGCTACTTACGGCTTAGCCGAGAGATATGCGG 1101
Db |||||
QY 925 GCTAAATATCTCTATGAGGACGCCAAGGTTTACCTACGAGCTCGAAGAGGATCTTCCCG 984
QY 1102 TTGTCTATCCAGTTGCTACGTAACAGGTCTCCCACTAGACCAGGTAGGTGCGATGAT 1161
Db |||||
QY 985 ATGAGCGGCTTGAGCGGCCAGAG---AGACTTACCGTGGCGCTATA 1266
QY 1162 GTTGGCTTTTCGACTTGAATGGTACTGTATACGCGCGCTTTAAAGATGAAGAGCTTGTG 1221
Db |||||
QY 1045 ACTGGCAACCTCGTCAGTGTCTCTCCGAGAGGCTATGAGAGGAAACGAGCTGGCC 1104
QY 1222 CCGAACCGGTTGAGCGGCCAGAG-----AGACTTACCGTGGCGCTATA 1266
Db |||||
QY 1105 CCGAAACAGCCGATGAAGAGGAGTGGCCAGAGACGCGCAGAGCTATGAAGGGGCTAC 1164
QY 1267 GTTCTTGAGCGTTGAGAGCGGTGCACAGAGATATAGCCGTACTCGACTTTAGCTCGATG 1326
Db |||||
QY 1165 GTAAAGAGCCGAGAGAGGCTATGGAAGAAACATAGTTTATCGATTTCCGTTCCGCTG 1224
QY 1327 TACCCAAACATCATATAAGTACATATGTTGGTCTCTGACACGCTTGTGAGGCTCGTGAA 1386
Db |||||
QY 1225 TACCCCTCAATCATCATACCCACACAGCTCTCGCGGATCTCTTAAAGGGAAGATGC 1284
QY 1387 AAGTGTGGCGAGTGTGGTGTGGGAGGCGCGGAGGTCAAGCAGAGTTTCCGTAGGTTG 1446
Db |||||
QY 1285 AAGGAGTATGACACT-----CCCCCGAGTTGGCCACCGCTTCTGCAAGGAC 1332
QY 1447 CCGCCCGGCTTCTCAAGACAGTTCTTGAGAGGCTGTTAGAGCTTCTGAAGCGTGTGCT 1506
Db |||||
QY 1333 TTCCCGGCTTCAATCCGAGCTCTTTGGGCGCTCTGTTGAGGAGGAGGAGAA---GATA 1389
QY 1507 GCTGAATGAAGAGTATCTCCGATAGCCAGAGATATCGACTGTGGATGAAGGCGAG 1566
Db |||||
QY 1390 AAGAAGAGATGAAGGCAAGATTGATCCACTTGAGAGAAAGCTCTTGTATACAGACAG 1449
QY 1567 AAGCGGTTGAAGGTTCTTGCAACCGTAGTTACGGCTACATGGGTTGAGCGCGCTAGG 1626
Db |||||
QY 1450 CCGGCTATCAAGATCCTAGCAACAGTTTATCGGCTACTACGGCTATCCAAAGGCCGCG 1509
QY 1627 TGGTATTGAGGAGTGCACAAAGCTGTCAGGCTTGGGTAGGCACTCATACGCAACC 1686
Db |||||
QY 1510 TGGTACTGCAAGAGTGTGAGAGAGCGTCACTGCTGGGGAAGGAGTACATACAGATG 1569
QY 1687 GCCATCA---ACATAGCTGTAACTAGGCTCAAGAGTATCTAGGTTGACACAGATTCG 1743
Db |||||
QY 1570 ACCATCAGGAGATGAGGAAAGTACGGCTTTAAAGTGTCTACGCTGTATACAGCCGT 1629
QY 1744 CTCTTCGTAACCTATGATCCGAGAGAGGTGGAATAATTCATCAAAATATTAAGGAGAG 1803
Db |||||
QY 1630 TTCTTCGCAACAATACCTGGAGCCGATGCTGAAACCGTCAAAAGAGGCAATGAGTTC 1689
QY 1804 CT-----GGGTTTCAAAATCAAGCTAGAGAGGTGTAC 1836
Db |||||
QY 1690 ATCAATATACATCAACAGAGCTTCCAGGTGCGTGTAGCTCGAGTATGAGGGTTTCTAC 1749
QY 1837 AAACGCTTATTTTACAGAGGCTAAGAGAGGTACGCTGGCTTCTCGAGGACGAGCT 1896

Db |||||
QY 1750 AAACCGGCTTCTT---CGTCAAGAAAGAGTACCGGTTGATAGACGAGAAAGGCAAG 1806
QY 1897 ATAGATATTTGCTGGTTTCGAGGCTGTACGTGGCGATTGGTGTGAACCTCGCAAGAGGTT 1956
Db |||||
QY 1807 ATAAACACGCGCGGACTTGTAGATTGTGAGCGCGACTGGAGCGAGATAGCGAAGAGACG 1866
QY 1857 CAGACTAAGCTTGTGCAAAATAGTATTGAAGACGAGTGAAGTGAACAAGGCTGTAGAGTAC 2016
Db |||||
QY 1867 CAGGCAAGGGTTCTTGAAGCTTTGCTAAAGACGCTGACGCTCGAAGAGCGCGTGAAGATA 1926
QY 2017 GTCAGGAAGATTGTGAAGAGTTGAGGAGGCGCAAGTTCCCATAGAGAGCTTGTATATC 2076
Db |||||
QY 1927 GTCAGGAGTCACTGAAAGCTGAGCAAGTACGAGGTTTCCCGGAGAGCTGGTGTATC 1986
QY 2077 TGGAGAGCCCTTAGTAAGCTCTTTAGAGGTACACAAGCGAGGCAACACAGCTCGTTGCA 2136
Db |||||
QY 1987 CACGAGCAGATAACGAGGATTTAAAGGACTTACAAGGCGACGGGACCCAGCTTCCGTT 2046
QY 2137 GCGAGAGGATGCTGTACAGAGGCTACCGGTTAAGCCAGCGCAAGAGTATGTA 2196
Db |||||
QY 2047 GCCAAGAGGTTAGCCGCGAGAGAGTCAAAATACGCCCTTGGACGCTGATAAGTACATC 2106
QY 2197 ATAGTGAAGGTTGTCGCGCTATCAGTCAAGAGGCTATGCGCATCTTCACTGCTCAAGAT 2256
Db |||||
QY 2107 GTGCTAAAGGCTCTGGAAGGATAGGCGACGAGGCAATACCGTTTCGACGAGTTCCGACCG 2166
QY 2257 CCTAGCCAGATAGACGTGACG---TACTATGTTGACCAACCAATATCATCCGCGCTGCATG 2313
Db |||||
QY 2167 ACGAAGCACAAGTACGACGCGAGTACTACATTGAGAACCAGGTTTCTCCAGCCGTTGAG 2226
QY 2314 AGRATCTGGGCTACTTTGGCATCCCGAGAGAGGCTG 2352
Db |||||
QY 2227 AGAATCTGAGAGCTTTCGGTTATFCGAAGGAGACCTG 2265

Search completed: November 26, 2003, 12:25:11

Job time : 635 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2003, 12:25:14 ; Search time 734 Seconds
(without alignments)
10826.693 Million cell updates/sec

Title: US-10-034-849-1
Perfect score: 2412
Sequence: 1 atgactgaagtgtattcac.....tagccaagaagaagtaaa 2412

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2190069 seqs, 1647345023 residues
Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2412	100.0	2412	9	US-09-391-340-3
2	2412	100.0	2412	10	US-09-948-369-3
3	2412	100.0	2412	13	US-09-948-369-13
4	2412	100.0	2412	13	US-10-034-849-1
5	2412	100.0	2412	13	US-10-034-621-1
6	2412	100.0	2412	13	US-10-029-382-1
7	2400.8	99.5	2412	10	US-09-948-369-15
8	279	11.6	2367	9	US-09-391-340-5
9	279	11.6	2367	10	US-09-948-369-5
10	190	7.9	2325	12	US-10-227-110-100
11	190	7.9	2325	12	US-10-208-508-100
12	189.8	7.9	2325	9	US-09-852-922-3
13	189.8	7.9	2325	12	US-10-227-110-97
14	189.8	7.9	2325	12	US-10-208-508-97
15	188.4	7.8	2325	12	US-10-227-110-98
16	188.4	7.8	2325	12	US-10-227-110-99

17	188.4	7.8	2325	12	US-10-227-110-101	Sequence 101, App
18	188.4	7.8	2325	12	US-10-227-110-102	Sequence 102, App
19	188.4	7.8	2325	12	US-10-227-110-103	Sequence 103, App
20	188.4	7.8	2325	12	US-10-208-508-98	Sequence 98, Appl
21	188.4	7.8	2325	12	US-10-208-508-99	Sequence 99, Appl
22	188.4	7.8	2325	12	US-10-208-508-101	Sequence 101, App
23	188.4	7.8	2325	12	US-10-208-508-102	Sequence 102, App
24	188.4	7.8	2325	12	US-10-208-508-103	Sequence 103, App
25	186.8	7.7	2325	12	US-10-227-110-104	Sequence 104, App
26	186.8	7.7	2325	12	US-10-208-508-104	Sequence 104, App
27	184.8	7.7	2325	9	US-09-803-165-33	Sequence 33, Appl
28	174.8	7.2	2331	12	US-09-896-923-1	Sequence 1, Appl
29	163.4	6.8	2322	12	US-10-227-110-123	Sequence 123, App
30	163.4	6.8	2322	12	US-10-208-508-123	Sequence 123, App
31	161.8	6.7	2322	12	US-10-227-110-122	Sequence 122, App
32	161.8	6.7	2322	12	US-10-208-508-122	Sequence 122, App
33	160.2	6.6	2322	12	US-10-227-110-121	Sequence 121, App
34	160.2	6.6	2322	12	US-10-208-508-121	Sequence 121, App
35	160.2	6.6	2322	12	US-10-227-110-125	Sequence 125, App
36	160.2	6.6	2322	12	US-10-208-508-125	Sequence 125, App
37	160.2	6.6	2322	12	US-10-227-110-126	Sequence 126, App
38	160.2	6.6	2322	12	US-10-208-508-126	Sequence 126, App
39	160.2	6.6	2322	12	US-10-208-508-124	Sequence 124, App
40	160.2	6.6	2322	12	US-10-208-508-125	Sequence 125, App
41	160.2	6.6	2322	12	US-10-208-508-126	Sequence 126, App
42	160.2	6.6	2322	12	US-10-208-508-127	Sequence 127, App
43	153	6.3	2325	12	US-10-227-110-108	Sequence 108, App
44	153	6.3	2325	12	US-10-208-508-108	Sequence 108, App
45	152.8	6.3	2325	12	US-10-227-110-105	Sequence 105, App

ALIGNMENTS

RESULT 1

US-09-391-340-3
; Sequence 3, Application US/09391340A
; Patent No. US20020013455A1
; GENERAL INFORMATION:
; APPLICANT: Callen, Walter
; APPLICANT: Mather, Eric
; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
; FILE REFERENCE: 09010/027001
; CURRENT APPLICATION NUMBER: US/09/391,340A
; CURRENT FILING DATE: 1999-09-07
; EARLIER APPLICATION NUMBER: US 08/907,166
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 2412
; TYPE: DNA
; ORGANISM: Pyrolobus fumarius
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2410)
US-09-391-340-3

Query Match	100.0%	Score 2412;	DB 9;	Length 2412;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 2412;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	ATGACTGAAGTTGTTATTCACGGTTTATAGCTCTAGCTACGAGGTTTGGTAAAGAGCCT	60	
DB	1	ATGACTGAAGTTGTTATTCACGGTTTATAGCTCTAGCTACGAGGTTTGGTAAAGAGCCT	60	
QY	61	CAGGTAATCATATCGGGTATTCCTGAGACGCGAGAGGTTAGTCTCTCTTGCAGGTTCT	120	
DB	61	CAGGTAATCATATCGGGTATTCCTGAGACGCGAGAGGTTAGTCTCTCTTGCAGGTTCT	120	
QY	121	TTTCGCCCATACTTCTATCGCTGCTTGCACGGGGCGGCATCCTTAAGCAGGTACACAA	180	
DB	121	TTTCGCCCATACTTCTATCGCTGCTTGCACGGGGCGGCATCCTTAAGCAGGTACACAA	180	

Db 2341 GAGAGAGCTGAAAGCAAGTGCACAACTGGGCGAGAGACTCTCTTCGACTTTTCTAGCCAAG 2400
 QY 2401 AAGAGCAAGTAA 2412
 Db 2401 AAGAGCAAGTAA 2412

RESULT 2

US-09-948-369-3
 ; Sequence 3, Application US/09948369
 ; Patent No. US20020132243A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DIVERSA CORPORATION
 ; APPLICANT: CALLEN, Walter
 ; APPLICANT: MATHUR, Eric
 ; TITLE OF INVENTION: ENZYMES HAVING HIGH TEMPERATURE POLYMERASE ACTIVITY AND METHODS
 ; TITLE OF INVENTION: THERMOF
 ; FILE REFERENCE: DIVER1350-3
 ; CURRENT APPLICATION NUMBER: US/09/948,369
 ; CURRENT FILING DATE: 2001-09-06
 ; PRIOR APPLICATION NUMBER: US 09/656,309
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: US 09/391,340
 ; PRIOR FILING DATE: 1999-09-07
 ; PRIOR APPLICATION NUMBER: US 08/907,166
 ; PRIOR FILING DATE: 1997-08-06
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: Patent in version 3.0
 ; SEQ ID NO 3
 ; LENGTH: 2412
 ; TYPE: DNA
 ; ORGANISM: Pyrololobus fumarius
 ; US-09-948-369-3

Query Match 100.0%; Score 2412; DB 10; Length 2412;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGACTGAAGTGTATTACGGTTTATAGACTCTAGCTACGAGGTTGTTGGTAAAGAGCT 60
 Db 1 ATGACTGAAGTGTATTACGGTTTATAGACTCTAGCTACGAGGTTGTTGGTAAAGAGCT 60
 QY 61 CAGGTAATCATATGGGTTATGCTGAGAACGGCGAGAGGTTAGTCTCTCATTCACAGGTC 120
 Db 61 CAGGTAATCATATGGGTTATGCTGAGAACGGCGAGAGGTTAGTCTCTCATTCACAGGTC 120
 QY 121 TTTCCGCCATCTCTATGGCTGCTTGCACCGGGGGCCGATCCCTAAGCAGGTAGCACAA 180
 Db 121 TTTCCGCCATCTCTATGGCTGCTTGCACCGGGGGCCGATCCCTAAGCAGGTAGCACAA 180
 QY 181 CGTATTCGTCATTGAGTAGGCGCAAGAGCCCGATTTATAGGTGTAGAGGATGACAAAGAG 240
 Db 181 CGTATTCGTCATTGAGTAGGCGCAAGAGCCCGATTTATAGGTGTAGAGGATGACAAAGAG 240
 QY 241 AAGTACTTGGGAGGCTCTAGGCTTACGTATTCGACCGCTGCTACCGAGGCTGTT 300
 Db 241 AAGTACTTGGGAGGCTCTAGGCTTACGTATTCGACCGCTGCTACCGAGGCTGTT 300
 QY 301 AGGAGTATCGCAACTCGTAAAGACGTTGATGTTGAGGATGTTCTAGAGCGGAT 360
 Db 301 AGGAGTATCGCAACTCGTAAAGACGTTGATGTTGAGGATGTTCTAGAGCGGAT 360
 QY 361 ATACGCTTCGCTATGCGCTATCTCATAGATCAAGTATCTATTTCTTCCCTGAGGCTGTT 420
 Db 361 ATACGCTTCGCTATGCGCTATCTCATAGATCAAGTATCTATTTCTTCCCTGAGGCTGTT 420
 QY 421 GTAGAGGCTGAGCCCTCGAGAACAGATGGGCTTCGGTTCGACAGGTATACCTGTT 480
 Db 421 GTAGAGGCTGAGCCCTCGAGAACAGATGGGCTTCGGTTCGACAGGTATACCTGTT 480
 QY 481 AAGAGCAGGCGGAGGCGGCTTTATGTTGAGGCTCTCGCACCAAGCAAGCTTCCCGATCTT 540

Db 481 AAGAGCAGGCGGAGGCGGCTTTATGTTGAGGCTCTCGCACCAACCAAGCTTCCCGATCTT 540
 QY 541 AAGATATCTCGGCTTCGATATTGAAGTTTATAGCAAGCAAGGTCGTCGCTCCAGAGCGC 600
 Db 541 AAGATATCTCGGCTTCGATATTGAAGTTTATAGCAAGCAAGGTCGTCGCTCCAGAGCGC 600
 QY 601 GATCCTGTAATAGTATAGCTGTGAGACTGACGATGCGGATGCGGATGCGTATCTCATTTGCA 660
 Db 601 GATCCTGTAATAGTATAGCTGTGAGACTGACGATGCGGATGCGGATGCGTATCTCATTTGCA 660
 QY 661 GAGGCGCAAGACGATCGAAACCGATACGCGAGTTTGTAGAGTAGTACGTTGAGAGGATGAC 720
 Db 661 GAGGCGCAAGACGATCGAAACCGATACGCGAGTTTGTAGAGTAGTACGTTGAGAGGATGAC 720
 QY 721 CCGGACATATAGTCTGGTTATACCAATCATTTTCGATTTGGCTTATCTTTGAGGCGC 780
 Db 721 CCGGACATATAGTCTGGTTATACCAATCATTTTCGATTTGGCTTATCTTTGAGGCGC 780
 QY 781 GCCCGCATCTTAGGCATAAAGCTTCATGTGACTAGAGAGTTGGCGCGAGCCACCACT 840
 Db 781 GCCCGCATCTTAGGCATAAAGCTTCATGTGACTAGAGAGTTGGCGCGAGCCACCACT 840
 QY 841 AGCTATATGCGGACGCTCTCTGCTCGGAGGCTTAAAGTAGTCTGTAGCACTATGCGC 900
 Db 841 AGCTATATGCGGACGCTCTCTGCTCGGAGGCTTAAAGTAGTCTGTAGCACTATGCGC 900
 QY 901 GAAGATGCGCAGAGATCAAGATTAAGAGTCTCGAGGAGTTCGAGAGTATCTAGGCGT 960
 Db 901 GAAGATGCGCAGAGATCAAGATTAAGAGTCTCGAGGAGTTCGAGAGTATCTAGGCGT 960
 QY 961 ATGAAGAAGAGTGAACGGCTTATCATTAATGTTGGGAGATTCGAGTATTTGGAGCAAC 1020
 Db 961 ATGAAGAAGAGTGAACGGCTTATCATTAATGTTGGGAGATTCGAGTATTTGGAGCAAC 1020
 QY 1021 CCGAAGAAGAGGACCACTATTACTCAATACGCGCGGAGTCTCCGCTACTTACGCGC 1080
 Db 1021 CCGAAGAAGAGGACCACTATTACTCAATACGCGCGGAGTCTCCGCTACTTACGCGC 1080
 QY 1081 TTAGCCGAGAAGATATTGCCGTTTGTCTATCCAGTTTGTCTAGTAACAGGCTCTCCCACTA 1140
 Db 1081 TTAGCCGAGAAGATATTGCCGTTTGTCTATCCAGTTTGTCTAGTAACAGGCTCTCCCACTA 1140
 QY 1141 GACAGGTAGTTCGATAGTGTGCTTTCGATTTGAATGTTACTGATACGCGCGCGC 1200
 Db 1141 GACAGGTAGTTCGATAGTGTGCTTTCGATTTGAATGTTACTGATACGCGCGCGC 1200
 QY 1201 TTTAAGATCAAGAGAGCTTGTGCGAACCGCTTGAGCGCCGAGAGAGACTTACCGTGC 1260
 Db 1201 TTTAAGATCAAGAGAGCTTGTGCGAACCGCTTGAGCGCCGAGAGAGACTTACCGTGC 1260
 QY 1261 GCTATAGTCTTTGAGCGGTTGAGAGCGGTCGACGAGATATAGACGCTTACGCTTACG 1320
 Db 1261 GCTATAGTCTTTGAGCGGTTGAGAGCGGTCGACGAGATATAGACGCTTACGCTTACG 1320
 QY 1321 TCCATATCCCAACATCATGATTAAGTACATGTTGCTGCTGACACGCTTGTGAGGCT 1380
 Db 1321 TCCATATCCCAACATCATGATTAAGTACATGTTGCTGCTGACACGCTTGTGAGGCT 1380
 QY 1381 GGTGAAAAGTGTGGCGAGTGTGCTGGGAGGCGCCGAGGCTCAAGCACAGGTTCCGT 1440
 Db 1381 GGTGAAAAGTGTGGCGAGTGTGCTGGGAGGCGCCGAGGCTCAAGCACAGGTTCCGT 1440
 QY 1441 AGGTGCGCGCGCTTCTTCAAGACAGTCTTGAAGAGCTGTTGAGAGCTTCTGTAAGGCT 1500
 Db 1441 AGGTGCGCGCGCTTCTTCAAGACAGTCTTGAAGAGCTGTTGAGAGCTTCTGTAAGGCT 1500
 QY 1501 GTCGCTGCTGAATGAAGAGTATCTCCGAGTACGAGATATCGACTGTGATGATGAA 1560
 Db 1501 GTCGCTGCTGAATGAAGAGTATCTCCGAGTACGAGATATCGACTGTGATGATGAA 1560
 QY 1561 AGGAGAGAGGCTTGAAGGTTCTTGAACCGTGTGAGGCTTACGCTACATGGGTTGGAGCGC 1620
 Db 1561 AGGAGAGAGGCTTGAAGGTTCTTGAACCGTGTGAGGCTTACGCTACATGGGTTGGAGCGC 1620

Db 841 AGCGTACATGGGACAGCTCTCTGTCCTCGCAGGCTTAACGTAGATCTGTACGACTATGCC 900
Qy 901 GAAGAGATGCCAGAGATCAAGATAAAGAGTCTCGAGAGGTGCGAGAGTATCTACGCGTG 960
Db 901 GAAGAGATGCCAGAGATCAAGATAAAGAGTCTCGAGAGGTGCGAGAGTATCTACGCGTG 960
Qy 961 ATGAAGAGAGTGAACGCGTATCATCAATTTGGTGGAGATTCAGAGACTATTTGGAGCGAC 1020
Db 961 ATGAAGAGAGTGAACGCGTATCATCAATTTGGTGGAGATTCAGAGACTATTTGGAGCGAC 1020
Qy 1021 CCGAGAGAGAGACCACTATTAAGAGTGAACGCGGAGAGAGTGTCCGCTACTTACGCGC 1080
Db 1021 CCGAGAGAGAGACCACTATTAAGAGTGAACGCGGAGAGAGTGTCCGCTACTTACGCGC 1080
Qy 1081 TTAGCCGAGAGATATTCGCGTTTGGCTTATCCAGTGTCTGTAACAGGTCTCCCACTA 1140
Db 1081 TTAGCCGAGAGATATTCGCGTTTGGCTTATCCAGTGTCTGTAACAGGTCTCCCACTA 1140
Qy 1141 GACCAGATAGGTGGATAGTGTGGCTTTGCACTTGAATGTGATCTGATAGCGCGCG 1200
Db 1141 GACCAGATAGGTGGATAGTGTGGCTTTGCACTTGAATGTGATCTGATAGCGCGCG 1200
Qy 1201 TTTAAGATGAAGAGCTTGTGCGGAAACCGGTTTGAAGCGCCAGAGAGACTTACCGTGC 1260
Db 1201 TTTAAGATGAAGAGCTTGTGCGGAAACCGGTTTGAAGCGCCAGAGAGACTTACCGTGC 1260
Qy 1261 GCTATAGTCTTGAAGCGCTTGAAGCGGTGCAAGAGATATAGCGGTACTCGACTTTAGC 1320
Db 1261 GCTATAGTCTTGAAGCGCTTGAAGCGGTGCAAGAGATATAGCGGTACTCGACTTTAGC 1320
Qy 1321 TCGATGTACCCAACTATGATTAAGATCAATGTTGCTTGAAGCGGTGATAGAGGCT 1380
Db 1321 TCGATGTACCCAACTATGATTAAGATCAATGTTGCTTGAAGCGGTGATAGAGGCT 1380
Qy 1381 GGTGAAGAGTGTGCGAGTGTGCTGCGAGGCGCCCGAGGTCAAGCACAGGTTCGT 1440
Db 1381 GGTGAAGAGTGTGCGAGTGTGCTGCGAGGCGCCCGAGGTCAAGCACAGGTTCGT 1440
Qy 1441 AGGTGCGCGCGCTTCTTCAAGACAGTCTTGAAGCGGTGATAGAGGCTTGAAGGT 1500
Db 1441 AGGTGCGCGCGCTTCTTCAAGACAGTCTTGAAGCGGTGATAGAGGCTTGAAGGT 1500
Qy 1501 GTGCGTCTGAATGAAGAGTATCTTCAAGAGTATCTTCAAGAGTATCTTCAAGAGTAA 1560
Db 1501 GTGCGTCTGAATGAAGAGTATCTTCAAGAGTATCTTCAAGAGTATCTTCAAGAGTAA 1560
Qy 1561 AGGCAGAGCGGTGAAGGTCTTGAAGAGTATCTTCAAGAGTATCTTCAAGAGTAA 1620
Db 1561 AGGCAGAGCGGTGAAGGTCTTGAAGAGTATCTTCAAGAGTATCTTCAAGAGTAA 1620
Qy 1621 GCTAGTGTGATTCAGAGGAGTGTGCGAAGGCTGTCAAGGCTTGGGTTAGGACCTCAT 1680
Db 1621 GCTAGTGTGATTCAGAGGAGTGTGCGAAGGCTGTCAAGGCTTGGGTTAGGACCTCAT 1680
Qy 1681 GCGACCGCATCAACATAGTCTGTAAGTCTAGGCTCAAGGTATCTACGCTGACAGAT 1740
Db 1681 GCGACCGCATCAACATAGTCTGTAAGTCTAGGCTCAAGGTATCTACGCTGACAGAT 1740
Qy 1741 TCGCTCTTCGTGACTATGATCCGAGAGGAGTGTGAAGTGTCAATTAATTAAGAGG 1800
Db 1741 TCGCTCTTCGTGACTATGATCCGAGAGGAGTGTGAAGTGTCAATTAATTAAGAGG 1800
Qy 1801 GAGCTGGGTTTGAAGTCAAGCTAGAGAGGTTGACAAAGCTTATCTTACAGAGGCT 1860
Db 1801 GAGCTGGGTTTGAAGTCAAGCTAGAGAGGTTGACAAAGCTTATCTTACAGAGGCT 1860
Qy 1861 AAGAGAGGTACGCTGCGCTTCTCGAGAGCGGAGTATGATATGTCGTTTCGAGGCT 1920
Db 1861 AAGAGAGGTACGCTGCGCTTCTCGAGAGCGGAGTATGATATGTCGTTTCGAGGCT 1920
Qy 1921 GTAGTGGCGATTTGCTGTAAGTGTGCAAGAGGTTTCAGACTAGGTTGCGAAATAGTA 1980

Db 1921 GTAGTGGCGATTTGCTGTAAGTGTGCAAGAGGTTTCAGACTAAAGTTGTCGAAATAGTA 1980
Qy 1981 TTGAAGACGAGTGTGAGTGTGAACAGGCTGTAGAGTACGTGAGAGATTTGTGAAGAGTTG 2040
Db 1981 TTGAAGACGAGTGTGAGTGTGAACAGGCTGTAGAGTACGTGAGAGATTTGTGAAGAGTTG 2040
Qy 2041 GAGAGGGGAGGTTTCCCATAGAGAGCTTGTAACTGGAAGACCTTGTAGTACGCTT 2100
Db 2041 GAGAGGGGAGGTTTCCCATAGAGAGCTTGTAACTGGAAGACCTTGTAGTACGCTT 2100
Qy 2101 GAGAGTACAAACGAGGAGGACCAACAGCTGCTGAGCGAGAGGATGCTGTCAGAGGCT 2160
Db 2101 GAGAGTACAAACGAGGAGGACCAACAGCTGCTGAGCGAGAGGATGCTGTCAGAGGCT 2160
Qy 2161 TACCGGTAAAGCCGAGGCGAGAGATGAGGTATGTAATAGTGAAGGTTGTCGCTATC 2220
Db 2161 TACCGGTAAAGCCGAGGCGAGAGATGAGGTATGTAATAGTGAAGGTTGTCGCTATC 2220
Qy 2221 AGTCAAGAGCATGGCCATCTTCAATGTCAGAGTCTTCAAGGATCCCTAGCAGATAGACCTAC 2280
Db 2221 AGTCAAGAGCATGGCCATCTTCAATGTCAGAGTCTTCAAGGATCCCTAGCAGATAGACCTAC 2280
Qy 2281 TATGTTGACCAACCAATCATCCCGCTGCAATGAGAAATCTGGGCTACTTTGGCATCAC 2340
Db 2281 TATGTTGACCAACCAATCATCCCGCTGCAATGAGAAATCTGGGCTACTTTGGCATCAC 2340
Qy 2341 GAGAGAGCTGGAAGCAGTGTCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGT 2400
Db 2341 GAGAGAGCTGGAAGCAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGT 2400
Qy 2401 AAGAGCAAGTAA 2412
Db 2401 AAGAGCAAGTAA 2412

RESULT 4

US-10-034-849-1
; Sequence 1, Application US/10034849
; Publication No. US20020115108A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: CALLEN, Walter
; APPLICANT: MATHUR, Eric
; TITLE OF INVENTION: ENZYMES HAVING HIGH TEMPERATURE POLYMERASE ACTIVITY
; TITLE OF INVENTION: AND METHODS OF USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: DIVER1350-2
; CURRENT APPLICATION NUMBER: US/10/034,849
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US/09/656,309
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2412
; TYPE: DNA
; ORGANISM: Pyrobolus fumarius
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2412)
US-10-034-849-1

Query Match 100.0%; Score 2412; DB 13; Length 2412;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGACTGAGTGTGATTTACGCTTTAGACTCTAGCTAGCTGAGGTTCTTGTGTAAGAGCT 60
Db 1 ATGACTGAGTGTGATTTACGCTTTAGACTCTAGCTAGCTGAGGTTCTTGTGTAAGAGCT 60
Qy 61 CAGTAAATCATATGCGGTATTTGTCGAGAGCGGAGAGGTTAGTCTCTTATGACAGGCT 120
Db 61 CAGTAAATCATATGCGGTATTTGTCGAGAGCGGAGAGGTTAGTCTCTTATGACAGGCT 120

121 TTTCCGCCATATCTTCTATGCGCTGCTTGCACCGGCGCCGATCTTAAGCGGTAGACAA 180
121 TTTCCGCCATATCTTCTATGCGCTGCTTGCACCGGCGCCGATCTTAAGCGGTAGACAA 180
181 CGTATTTCGTGCTAGTAGGCGCAAGAGCGCGATATTAGGTGTAGAGGATGACAAGG 240
181 CGTATTTCGTGCTAGTAGGCGCAAGAGCGCGATATTAGGTGTAGAGGATGACAAGG 240
241 AAGTACTTCGGAGCGCTCGTAGGCTTTACGTATTCGCAACGCTGCTACCGAGGCTGTT 300
241 AAGTACTTCGGAGCGCTCGTAGGCTTTACGTATTCGCAACGCTGCTACCGAGGCTGTT 300
301 AGGAGTATCGGCACTCGTAAAGAACGTTGATGCTGTAGAGGATGTTTAGAGCGCGAT 360
301 AGGAGTATCGGCACTCGTAAAGAACGTTGATGCTGTAGAGGATGTTTAGAGCGCGAT 360
361 ATACGCTTCGCTATGCGCTATCTCATAGATCACGATCTATTTCCTTTCACTGCTACCGT 420
361 ATACGCTTCGCTATGCGCTATCTCATAGATCACGATCTATTTCCTTTCACTGCTACCGT 420
421 GTAGAGCTGAGCGCTCGAGAACAGATGCGCTTCCGTCGCAACGCTGCTACCGT 480
421 GTAGAGCTGAGCGCTCGAGAACAGATGCGCTTCCGTCGCAACGCTGCTACCGT 480
481 AAGAGCAGCGCGAGCGCTTTATGCTGAGCTCTCGCAACGCTGCTACCGT 540
481 AAGAGCAGCGCGAGCGCTTTATGCTGAGCTCTCGCAACGCTGCTACCGT 540
541 AGGATACTCGCTTCGATATGAACTTATAGCAAGAGGTCGCTGCGCAACGCTGCTACCGT 600
541 AGGATACTCGCTTCGATATGAACTTATAGCAAGAGGTCGCTGCGCAACGCTGCTACCGT 600
601 GATCTGTAATAGTATGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 660
601 GATCTGTAATAGTATGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 660
661 GAGGCAAGAGCGTAAAGCGGATACGCGGATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 720
661 GAGGCAAGAGCGTAAAGCGGATACGCGGATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 720
721 CCGCATATATAGTGGTATTAACCAACATCATTTTCGATTCGCTTATCTTTTTCGAGCGC 780
721 CCGCATATATAGTGGTATTAACCAACATCATTTTCGATTCGCTTATCTTTTTCGAGCGC 780
781 GCGCGATCTTAGGCTAAAGCTTGTATGCTAGAGGTCGCGCGAGCTGAGCTGAGCTGAGCT 840
781 GCGCGATCTTAGGCTAAAGCTTGTATGCTAGAGGTCGCGCGAGCTGAGCTGAGCTGAGCT 840
841 AGCGTACATGGCGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
841 AGCGTACATGGCGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
901 GAAGAGATCCGAGAGATCAAGTAAAGAGTCTCGAGAGGTCGCGAGGATCTAGAGGCTG 960
901 GAAGAGATCCGAGAGATCAAGTAAAGAGTCTCGAGAGGTCGCGAGGATCTAGAGGCTG 960
961 ATGAGAGAGGTGAACGCTTATCATTAATGCTGAGAGTTCAGAGCTATTTGGAGCGAC 1020
961 ATGAGAGAGGTGAACGCTTATCATTAATGCTGAGAGTTCAGAGCTATTTGGAGCGAC 1020
1021 CCGAAGAGAGACCACTATTAATGCAATACGCGCGAGAGTTCGCGCTACTTACCGG 1080
1021 CCGAAGAGAGACCACTATTAATGCAATACGCGCGAGAGTTCGCGCTACTTACCGG 1080
1081 TTAGCGGAGAGATATTCGCTTCTCTATCCAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
1081 TTAGCGGAGAGATATTCGCTTCTCTATCCAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
1141 GACCAAGAGAGATATTCGCTTCTCTATCCAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
1141 GACCAAGAGAGATATTCGCTTCTCTATCCAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200

1201 TTTAAGATGAAGAGCTTGTGCGCAACCGGCTTGAAGCGCCCGAGAGAGCTTACCGTGGC 1260
1201 TTTAAGATGAAGAGCTTGTGCGCAACCGGCTTGAAGCGCCCGAGAGAGCTTACCGTGGC 1260
1261 GCTATAGTTCCTTGAAGCGCTTGAAGAGGCTGCAAGAGATATAGCGCTTACCGTGGC 1320
1261 GCTATAGTTCCTTGAAGCGCTTGAAGAGGCTGCAAGAGATATAGCGCTTACCGTGGC 1320
1321 TCGATGATACCAACATCATGATTAAGTCAATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
1321 TCGATGATACCAACATCATGATTAAGTCAATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
1381 GGTGAAAGTGTGCGGAGTGTGCTGCTGCGAGGCTGCGAGGCTCAAGACAGCTTCCCT 1440
1381 GGTGAAAGTGTGCGGAGTGTGCTGCTGCGAGGCTGCGAGGCTCAAGACAGCTTCCCT 1440
1441 AGGTGTCGCGCGGCTTCTTCAAGACAGCTTCTTGAAGGCTGTTAGAGCTTCGTAAGCGT 1500
1441 AGGTGTCGCGCGGCTTCTTCAAGACAGCTTCTTGAAGGCTGTTAGAGCTTCGTAAGCGT 1500
1501 GTGCGTGTGAAATGAAGAGTATCTTCCGAGAGGCTGCGAGGCTCAAGATGATGATGAA 1560
1501 GTGCGTGTGAAATGAAGAGTATCTTCCGAGAGGCTGCGAGGCTCAAGATGATGATGAA 1560
1561 AGGCAAGAGGCTTGAAGGCTTCTTGAAGGCTTCTTGAAGGCTTCTTGAAGGCTTCTTGAAGGCT 1620
1561 AGGCAAGAGGCTTGAAGGCTTCTTGAAGGCTTCTTGAAGGCTTCTTGAAGGCTTCTTGAAGGCT 1620
1621 GCTAGTGTGATTCGAGGAGTGTGCGAGGCTGCTGCGAGGCTTGAAGGCTTCTTGAAGGCT 1680
1621 GCTAGTGTGATTCGAGGAGTGTGCGAGGCTGCTGCGAGGCTTGAAGGCTTCTTGAAGGCT 1680
1681 GCGACCGGCTATCATAGCTGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCT 1740
1681 GCGACCGGCTATCATAGCTGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCT 1740
1741 TCGCTTCTGCTGAGCTATGATTCGAGAGGCTGCGAGGCTTGAAGGCTTGAAGGCTTGAAGGCT 1800
1741 TCGCTTCTGCTGAGCTATGATTCGAGAGGCTGCGAGGCTTGAAGGCTTGAAGGCTTGAAGGCT 1800
1801 GAGTGGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCT 1860
1801 GAGTGGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCTTGAAGGCT 1860
1861 AAGAGAGGCTGAGTGGCTTCTGCGAGGAGGCTGATGATGATGATGATGATGATGATGATGATGAT 1920
1861 AAGAGAGGCTGAGTGGCTTCTGCGAGGAGGCTGATGATGATGATGATGATGATGATGATGATGAT 1920
1921 GTACGTCGCGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
1921 GTACGTCGCGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
1981 TTGAGACGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2040
1981 TTGAGACGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2040
2041 GAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2100
2041 GAGGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2100
2101 GAGGAGTACACACGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2160
2101 GAGGAGTACACACGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2160
2161 TACCGGCTGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2220
2161 TACCGGCTGAGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2220
2221 AGTCAAGAGCATGCGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2280
2221 AGTCAAGAGCATGCGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2280
2281 TATGTTGACCAACATCATCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2340

QY 61 CAGGTAAATCATATGGGTAATGCTGAGAACGGCGAGAGGGTAGTCTCTCAATGACAGGTCT 120
DB 61 CAGGTAAATCATATGGGTAATGCTGAGAACGGCGAGAGGGTAGTCTCTCAATGACAGGTCT 120
QY 121 TTTCCGCCCATCTCTATGCGCTGCTTGACACCGGCGCGGATCTTAAGCAGGTAGCAAA 180
DB 121 TTTCCGCCCATCTCTATGCGCTGCTTGACACCGGCGCGGATCTTAAGCAGGTAGCAAA 180
QY 181 CGTATTCGTGCTGATGAGTAGCCAAAGAGCCCGGATTTATAGTGTAGAGGATGACAAGAGG 240
DB 181 CGTATTCGTGCTGATGAGTAGCCAAAGAGCCCGGATTTATAGTGTAGAGGATGACAAGAGG 240
QY 241 AAGTACTTCCGGAGGCTCTGTAAGGCTCTTACGTATCCACCGTGTACCCGAGGCTGTT 300
DB 241 AAGTACTTCCGGAGGCTCTGTAAGGCTCTTACGTATCCACCGTGTACCCGAGGCTGTT 300
QY 301 AGGAGATATCCGAACTCGTAAGACGTTGATGCTGTAGAGATGTTCTAGAGGCGGAT 360
DB 301 AGGAGATATCCGAACTCGTAAGACGTTGATGCTGTAGAGATGTTCTAGAGGCGGAT 360
QY 361 ATACGCTTCGTATGCGCTATCTCATAGATCACGATCTATTTCCTTTCACTGTACCGT 420
DB 361 ATACGCTTCGTATGCGCTATCTCATAGATCACGATCTATTTCCTTTCACTGTACCGT 420
QY 421 GTAGAGGCTGAGCCCTCGAGAACAGATGCGCTTCCGTGTGACAAAGGTATACCTGGTT 480
DB 421 GTAGAGGCTGAGCCCTCGAGAACAGATGCGCTTCCGTGTGACAAAGGTATACCTGGTT 480
QY 481 AAGAGAGGCGGAGGACCTTTATGCTGAGGCTCTCGACCAACCAAGCTTCCGATCTT 540
DB 481 AAGAGAGGCGGAGGACCTTTATGCTGAGGCTCTCGACCAACCAAGCTTCCGATCTT 540
QY 541 AGGATACTCGGCTTCGATATTGAAGTTTATAGCAAGCAAGGCTCGCGGCTCGAGAGCG 600
DB 541 AGGATACTCGGCTTCGATATTGAAGTTTATAGCAAGCAAGGCTCGCGGCTCGAGAGCG 600
QY 601 GATCCTGTAATAGTGTAGTGTGAGACTGACGATGCGGATGAGGTGCTATTCAATTGCA 660
DB 601 GATCCTGTAATAGTGTAGTGTGAGACTGACGATGCGGATGAGGTGCTATTCAATTGCA 660
QY 661 GAGGCAAGACAGATCGAAACCGATACGCGGCTTGTAGAGTACGCTGAGAGGATGAC 720
DB 661 GAGGCAAGACAGATCGAAACCGATACGCGGCTTGTAGAGTACGCTGAGAGGATGAC 720
QY 721 CCCGACATAATAGTCGGTTATTAACAACAATCAATTCGATTCGCTTATCTTTGAGGCG 780
DB 721 CCCGACATAATAGTCGGTTATTAACAACAATCAATTCGATTCGCTTATCTTTGAGGCG 780
QY 781 GCCCGCATCTAGGCTAAAGCTTGATGTGACTAGAGAGTTCGCGCCGAGCCCACT 840
DB 781 GCCCGCATCTAGGCTAAAGCTTGATGTGACTAGAGAGTTCGCGCCGAGCCCACT 840
QY 841 AGCGTACATGGGCACTCTGTCCTCGGAGGCTTAACTAGATCTGTACGACTATGCC 900
DB 841 AGCGTACATGGGCACTCTGTCCTCGGAGGCTTAACTAGATCTGTACGACTATGCC 900
QY 901 GAGAGATGCCAGATCAAGATAAAGAGTCTCGAGAGGTCCAGAGTATCTAGGCGTG 960
DB 901 GAGAGATGCCAGATCAAGATAAAGAGTCTCGAGAGGTCCAGAGTATCTAGGCGTG 960
QY 961 ATCAAGAGAGTGAACCGGTATCATCAATTTGTTGGGAGATTCAGACTATTGGGACGAC 1020
DB 961 ATCAAGAGAGTGAACCGGTATCATCAATTTGTTGGGAGATTCAGACTATTGGGACGAC 1020
QY 1021 CCGAAGAGAGACCACTATTACTGCAATACGCGGAGGATGTCGCGCTACTTACGCG 1080
DB 1021 CCGAAGAGAGACCACTATTACTGCAATACGCGGAGGATGTCGCGCTACTTACGCG 1080
QY 1081 TTAGCCGAGAGATATTGCGGTTTGTCTTCCAGTGTCTGCTAGTAAACAGGCTCTCCACTA 1140
DB 1081 TTAGCCGAGAGATATTGCGGTTTGTCTTCCAGTGTCTGCTAGTAAACAGGCTCTCCACTA 1140
QY 1141 GACCAGGTAGTCGATGAGTGTTCGCTTTGCAATTTGATGCTGATACGCGGCG 1200

DB 1141 GACCAGGTAGTCGATGAGTGTTCGCTTTTCAATTTGATGCTGATACGCGGCG 1200
QY 1201 TTTAAGATGAAGAGCTTGTGCCAAACCGCGTTTGAGCGCCCAAGAGAGACTTACCGTGCG 1260
DB 1201 TTTAAGATGAAGAGCTTGTGCCAAACCGCGTTTGAGCGCCCAAGAGAGACTTACCGTGCG 1260
QY 1261 GCTATAGTCTTGAGCCGTTGAGCGCGTGCACGAGATATATAGCCGTTACTCGACTTTAGC 1320
DB 1261 GCTATAGTCTTGAGCCGTTGAGCGCGTGCACGAGATATATAGCCGTTACTCGACTTTAGC 1320
QY 1321 TCGATGTACCCAAACATCATGATAAAGTACAAATGTTGGTCTTGACACCTCTGTGAGGCT 1380
DB 1321 TCGATGTACCCAAACATCATGATAAAGTACAAATGTTGGTCTTGACACCTCTGTGAGGCT 1380
QY 1381 GGTGAAAGTGTGGCGAGTGTGCTTGGGAGGCCCCCGAGGTTCAAGCACAGGTTCCGT 1440
DB 1381 GGTGAAAGTGTGGCGAGTGTGCTTGGGAGGCCCCCGAGGTTCAAGCACAGGTTCCGT 1440
QY 1441 AGGTGTCCGCCCGGCTTCTTCAAGACAGTCTTGTAGAGGCTGTTTAGAGCTTCGTAAGCGT 1500
DB 1441 AGGTGTCCGCCCGGCTTCTTCAAGACAGTCTTGTAGAGGCTGTTTAGAGCTTCGTAAGCGT 1500
QY 1501 GTGCGTCTGAAATCAAGAAAGTATCTCCGATAGCCAGAAATATCGACTGTTGATCAA 1560
DB 1501 GTGCGTCTGAAATCAAGAAAGTATCTCCGATAGCCAGAAATATCGACTGTTGATCAA 1560
QY 1561 AGGCAAGAGGCGCTTGAAGTCTTTCGAAACCGTAGTTACGGCTACATCGGTTGGAGCGCG 1620
DB 1561 AGGCAAGAGGCGCTTGAAGTCTTTCGAAACCGTAGTTACGGCTACATCGGTTGGAGCGCG 1620
QY 1621 GCTAGTGTGTAATGAGAGGCTGCGCAAGGCTGTACGGCTTGGGTTAGGACCTCATATA 1680
DB 1621 GCTAGTGTGTAATGAGAGGCTGCGCAAGGCTGTACGGCTTGGGTTAGGACCTCATATA 1680
QY 1681 CSCACCGGCTATCAACATAGCTCGTAAACTAGGCTTCAAGGCTGATCTAGGCTGACACAGAT 1740
DB 1681 CSCACCGGCTATCAACATAGCTCGTAAACTAGGCTTCAAGGCTGATCTAGGCTGACACAGAT 1740
QY 1741 TCGCTCTTCGTGACCTATGATCCGAGAGGCTGAAATTTTCATCAAAATATATAAGAG 1800
DB 1741 TCGCTCTTCGTGACCTATGATCCGAGAGGCTGAAATTTTCATCAAAATATATAAGAG 1800
QY 1801 GAGCTGGGTTGCGAATCAAGCTAGAGAGGCTGACAAACGCTTATCTTTTACAGAGCT 1860
DB 1801 GAGCTGGGTTGCGAATCAAGCTAGAGAGGCTGACAAACGCTTATCTTTTACAGAGCT 1860
QY 1861 AAGAGAGGTACGCTGCGCTTCTCGAGGACGACCTATAGATATTGTCGTTTCCGAGCT 1920
DB 1861 AAGAGAGGTACGCTGCGCTTCTCGAGGACGACCTATAGATATTGTCGTTTCCGAGCT 1920
QY 1921 GTACGTGCGATTTGTTGTAACCTCCCAAGAGGTTTCAAGTAAAGTTTGTGCGAATAGTA 1980
DB 1921 GTACGTGCGATTTGTTGTAACCTCCCAAGAGGTTTCAAGTAAAGTTTGTGCGAATAGTA 1980
QY 1981 TTGAAGACGAGTGTGAGTGAACAAGCTGTAGAGTACCTGAGAGGATTTGGAAGAGTTG 2040
DB 1981 TTGAAGACGAGTGTGAGTGAACAAGCTGTAGAGTACCTGAGAGGATTTGGAAGAGTTG 2040
QY 2041 GAGGAGGCAAGGTTCCCATAGAGAGCTTGTATCTGGAAGACCTTGTAGTAAAGCTT 2100
DB 2041 GAGGAGGCAAGGTTCCCATAGAGAGCTTGTATCTGGAAGACCTTGTAGTAAAGCTT 2100
QY 2101 GAGAGTATACAAACGAGGCTTCCACAGCTGCTGACGAGAGGATGCTGTGACAGGCG 2160
DB 2101 GAGAGTATACAAACGAGGCTTCCACAGCTGCTGACGAGAGGATGCTGTGACAGGCG 2160
QY 2161 TACCGGTGAGCCGAGGCTGAGATAGGCTTGTATAGTAAAGGCTGTGCGCGTATC 2220
DB 2161 TACCGGTGAGCCGAGGCTGAGATAGGCTTGTATAGTAAAGGCTGTGCGCGTATC 2220
QY 2221 AGTCAAGAGCATGGCCATCTTCTCATGCTCAAGGATCTTAGCCAGATAGACGCTAC 2280

Db 2221 AGTCAAAGAGCATGGCCATCTTCATGGTCAAGGATCCCTAGCCAGATAGACGTGACCTAC 2280
QY 2281 TATGTTGACCAACCAATCATCCCGCTGCAATGAGAAATACTGGGCTACTTTGGCATCACC 2340
Db 2281 TATGTTGACCAACCAATCATCCCGCTGCAATGAGAAATACTGGGCTACTTTGGCATCACC 2340
QY 2341 GAGAAGAGCTGAAGCAAGTGCACCTGGGCGAGAGACTCTCTTGGACTTTCTAGCCRAAG 2400
Db 2341 GAGAAGAGCTGAAGCAAGTGCACCTGGGCGAGAGACTCTCTTGGACTTTCTAGCCRAAG 2400
QY 2401 AAGGCAAGTAA 2412
Db 2401 AAGGCAAGTAA 2412

RESULT 8
US-09-391-340-5
; Sequence 5, Application US/09391340A
; Patent No. US20020013455A1
; GENERAL INFORMATION:
; APPLICANT: Callen, Walter
; APPLICANT: Mather, Eric
; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
; FILE REFERENCE: 09010/027001
; CURRENT APPLICATION NUMBER: US/09/391,340A
; EARLIER FILING DATE: 1999-09-07
; EARLIER APPLICATION NUMBER: US 08/907,166
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 2367
; TYPE: DNA
; ORGANISM: Archaeoglobus lithotrophicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2364)
US-09-391-340-5

Query Match 11.6%; Score 279; DB 9; Length 2367;
Best Local Similarity 50.9%; Pred. No. 4.1e-80;
Matches 889; Conservative 0; Mismatches 795; Indels 63; Gaps 7;

QY 530 TTCCCGATCTTAGGACTCGGCTTCGATATTGAAGTTTATAGCAAGCAAGGCTGCGCGC 589
Db 488 TTCCCGAATCAAGGTTATGCGGTTGATTGCGAATGCTCTCAGAGGTTGGTATGCGCG 547
QY 590 GTCCAGAGCGGCTCTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTG 649
Db 548 ATCCAGAGAGATCCATCATAGTATCATATCAATTAATTAATTAATTAATTAATTAAT 607
QY 650 TATTCAATTCAGAGGCAAGACGATCGAAACCCGATACGCGAGTTTGTAGAGTACGTGA 709
Db 608 TCAACGGTG-----ATAAGAGAGAGATGCTTACCAGATTGTCAAGATATTC 658
QY 710 AGAGTATGACCCGACATATAGTGGTTATACAAATCATTTGATTTGGCTTATC 769
Db 659 GCGATATTGATCCGACATTAAGTTGGATACATACAGCAGCTTTGACTGGCCCTATA 718
QY 770 TTTTGGGCGCGCCGATCCCTAGGCAATAGCTTGTGTAAGTGTGTAAGTGTGTAAGTGT 829
Db 719 TCAAGAGAGAGCTGAGAACTGAGGTTTAAAGCTTGAATCGATCGAAGAGATAGAAAGAAC 778
QY 830 AGCCCACTAGCTAGCTACATGGCGCTCTGTCCTGCGCAGGCTTAACGTAGATCTGT 889
Db 779 TGCTATCAGGCGAGAGACCAAA-----GATTGCTGCGCAGGTTGACCGTGTCT 832
QY 890 ACGACTATCCGAGAGATGCGCAGAGATCAAGATTAAGAGTCTCAGAGAGGTCGACAGT 949
Db 833 ATGATATTGCAATGAGGAGTCTCGATGTAAGGTCGAGAGCTCGAAACGTTGACAGT 892
QY 950 ATCTAGGCGGTGTAAGAGAGTGAACGCGTTATCATCAATTTGGTGGAGATTTCCAGACT 1009

Db 893 TTCTGGGTAGAAATAAGAGCTTCAGATATTGAAGCGAAGGATATCTTACAGCACTGGA 952
QY 1010 ATTGGGACGACCCGAGAGAGAGACCACTATTACTTGCAATACGCGCGGACGATGTCGCG 1069
Db 953 CATCGGGCGAGAGGAAAGCG-----TAATCAATATCTCCCGCGAGACATCTCTGC 1003
QY 1070 CTACTTAGCGGCTTAGCCGAGAGATATTGCGGTTTGTCTATCCAGTTGTCGTAACACAG 1129
Db 1004 ACAGCTACTTCATAGCTGAGAAATGCTGCCCAATGCAATTCGAACTTTCCGAATGATAC 1063
QY 1130 GTCTCCCACTAGACCAAGTAGGTGCGATGAGTGTGGCTTTTCGACTTTGAATGGTACCTGA 1189
Db 1064 GCATACCTCTCGATGATGTGACAAGGAGCGGAGAGGTAAAGAGTTGAGTGGCTGTGT 1123
QY 1190 TACGCGCGCGCTTTAAGATGAAGAGCTTTGCGCGAACCCTGTTGAGCGCCGAGAAAGA 1249
Db 1124 TAAGCAAGCACACAAACTTTGCGCAACTTGCACCCCAACCCCGAGAGAG---ATGCCGACA 1180
QY 1250 CTTACCGTGGCGCTATAGTTCTTTCAGCGGTTGAGAGCGTGCACGAGAAATATATATATAT 1309
Db 1181 GCTATGAGAGAGCAATTCGTGCTCGAGCCCGCAAGAGGATTCATGAGAACTATATCTGCC 1240
QY 1310 TCGACTTTAGCTCGATGTACCCAAACATCATGATAAAGTACAAATGTTGGTCTCTGACACGC 1369
Db 1241 TCGACTTTGCGTCCATGTATCCCTCAATTAATGATTTTCATACACATCAGCCCGACACGC 1300
QY 1370 TTGTGAGCGCTGTTGAAAGTGTGCGGAGTGTGTTGCTGGGAGGCGCCGAGGTCGAAGC 1429
Db 1301 TTGT-----AATAGGCAATTCGCGAGGATTCGAATGTAGCCCGGAGTGGGGC 1348
QY 1430 ACAGGTTCCGTTAGGTTGCGCGCGCTTCTTCAAGACAGTCTTTCGAGAGGCTGTAGAGC 1489
Db 1349 ACAATTCAGGAAACATCTCTGATGTTTTCGAAAGAACTATCAAAATGCTGATGAGA 1408
QY 1490 TTCGTAAGCGGTGCGGTGTAATGAAGAGTATCTCCGAGTAGCCGAGAAATATATCGAC 1549
Db 1409 AAAGAGAGAAATAAGAGGTTATGAACACATTCGACTACAACTCGCCGAGATATCAAGC 1468
QY 1550 TGTGGATGAAGGCAAGCGGTTGAGGTTCTTTCGAAAGCTTCTGCAAGCTGTAGCGCTACATGG 1609
Db 1469 TGCTCGATATAAGACGCGCAACGCTGAAAGTTCTTACAACTCGTTTACGGTTTATCTG 1528
QY 1610 GTTGGAGCGGCTGAGGTGTTATTCAGGAGTGCAGAAAGGCTGTCACCGCTTGGGGTA 1669
Db 1529 GGTGAGTCTTTCGAGATGTTACTGCAAGGAGTGCCTGAGCTACAAACGCTGCGGCA 1588
QY 1670 GGCACCTCTACGACCGCCATCAACATGCTGTAACCTAGGCTCAAGGCTGATCTACG 1729
Db 1589 GACACTTTATCAAAACATCTGCAAGAAATTCGAAAGAGCTTGGATTGGAAGTGTATATG 1648
QY 1730 GTGACACAGATTTCGCTCTGTCACCTATGATCCGAGAGAGTGGAAATTTTCATCAAAA 1789
Db 1649 GGGATACAGATAGCATCTTTGTTAAAGAGATGATGATTCGAGCTGGAAGAGCTCAAAAAG 1708
QY 1790 TTATAAA-----CGAGGAGCTGGGGTTTTCGAAATCAAGCTAGAGA 1828
Db 1709 AAGTTAAAGCTCATAGGTAAACTTTTCGAAAGAGATGCCATACAAATAGAGATATG 1768
QY 1829 AGGTGTCAACAGCTTATCTTTCAGAGGCTAAGAGAGTACGCTGGCTTCTCGAGG 1888
Db 1769 AATACTACGAGACAAATATCTTCTGTTGAA---AAGAAAGGTTATGCTGGATTGACACAG 1825
QY 1889 ACGGACCTATAGATATTGTCGGTTCGAGGCTGTAGTGGGATTTGGTGTGAATCTGCCA 1948
Db 1826 ATGGAAGATAATTTGTTAAAGGCTTCTGAACTCAGAGAGGCGGCTGCTGCGAGCTTGCMA 1885
QY 1949 AGGAGGTTTCAGACTAAGGTTGTCGAATAGTATTGGAAGAGAGTGAAGGTGAACAGGCTG 2008
Db 1886 AAGAGATACAGAGAGGTTGTAATAGAAATCATCTTGAAGGAAAGAAATCTTGAAGAAAGCTG 1945
QY 2009 TAGAGTACGTGAGAGAGATTTGAAAGAGTTGGAGGAGGCGAAGGTTTCCCATAGAGAGC 2068
Db 1946 CTGAGTATGTGAAAGGAGTCTAGAGGAGATTAAGGAGGCAAAATTCGCTTGAAGATT 2005

QY 2069 TTGTAATCTGGAAGACCCCTTAGTAAGCGCTCTTGAGAGGTACACAACGGAGGACCAACAGC 2128
Db 2006 ATATCATCTACAGAGGATTGACGAGAAACCATCAAAAGTACGAGATATGACGGCTCAGC 2065
QY 2129 TCGTTGACGCGAGAGAGGTGCTGTCAGCAGCTACCGGGTAAGCCAGGCGCAAGATAG 2188
Db 2066 TAAAGCTGCTGAGCGGCAAGAGAGGATAGTATACATCCGCTCAAGGTTG 2125
QY 2189 GGTATGTAATAGTAAGGCTGCTGCGCGTATCATGTCAAAGAGCATGSCCATCTTCATGG 2248
Db 2126 GTTTGCTGTTACAAAGGCTGCGGGAACATAGGTATAGGCTTTTCCATCTGATCTGA 2185
QY 2249 TCAAGGA 2255
Db 2186 TAGAGGA 2192

RESULT 9

US-09-948-369-5
; Sequence 5, Application US/09948369
; Patent No. US20020132243A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: CALLEN, Walter
; APPLICANT: MATHUR, Eric
; TITLE OF INVENTION: ENZYMES HAVING HIGH TEMPERATURE POLYMERASE ACTIVITY AND METHODS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: DIVER1350-3
; CURRENT APPLICATION NUMBER: US/09/948,369
; CURRENT FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: US/09/656,309
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US/09/391,340
; PRIOR FILING DATE: 1999-09-07
; PRIOR APPLICATION NUMBER: US/08/907,166
; PRIOR FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 2367
; TYPE: DNA
; ORGANISM: Archaeoglobus lithotrophicus
US-09-948-369-5

Query Match 11.6%; Score 279; DB 10; Length 2367;
Best Local Similarity 50.9%; Pred. No. 4.1e-80;
Matches 889; Conservative 0; Mismatches 795; Indels 63; Gaps 7;
QY 530 TTCCCGACTAGGATCTCGGTTGATATGAGTTTATAGCAAGGAGGTCGCGC 589
Db 488 TTCCCGAACTCAAGGTTATGCGGTTGATTGCGAAATGCTCTCAGAGGTTGGTATGCGC 547
QY 590 GTCCAGAGCGATCCTGTAATAGTATGATGCTGTGAGAGCTGACGATGCGGATGAGTGC 649
Db 548 ATCCAGAGAGACTGATCATATGTCATATCAATTAATCGGTGAAATACAGGAATCC 607
QY 650 TATTCATTGACAGAGGCGAAGACGATGCAAAACCGATACGCGAGTTTGTAGAGTACGTGA 709
Db 608 TCAACGGTG-----ATAACGAGAGAGATGCTTACAGATTGTCAGATATTC 658
QY 710 AGAGTATGACCCGACATATGCTGTTATACACATCATCTTCGATGCGCTTATC 769
Db 659 GCGATATTGATCCGACATATTATGTTGATACATCAGGACAGCTTTGACTGCGCCCTATA 718
QY 770 TTTTGAGCGCGCCCGCATCTTAGGCATATAAGCTTGATGTGACTAGAGAGGTTGGCGCCG 829
Db 719 TCAAGAGAGAGCTGAGAACTGAGGGTTAGCTTGATCTCGAGAGATAGAGCGAAC 778
QY 830 AGCCACCCNCTAGCTGATCATGGGACGCTCTCTGTCCTGCGAGGCTTAACGTAGATCTGT 889
Db 779 TGGCTATCAGGGGAGAGAACCAAA-----GATTGCTGGCAGGTTGAACTGATCTCT 832

QY 890 ACAGTATGCGAGAGATGCCAGAGATCAAGATAAGAGTCTTCGAGAGGCTCCAGAGT 949
Db 833 ATGATATTGCAATGAGAGTCTCGATGTAAGAGTGAAGAGCTCGAAGACGTTGCGAGT 892
QY 950 ATCTAGGCGTATCAAGAGAGTGAACGCTTATCATCAATTTGGTGGAGATTCAGAGCT 1009
Db 893 TTCTGGGTAAAGAAATAGACTTTCAGATATTAGCGAAGGATATCTACAGACTCGA 952
QY 1010 ATTGGAGAGCCCGAAGAGAGACCACTATTACTGCAATACGCGCGAGAGTGTCCGG 1069
Db 953 CATCGGCGAGCAGGAAAGCG-----TAATCAAAATACTCCCGCGAGAGATCTCTGC 1003
QY 1070 CTACTTACCGCTAGCGAGAGATATTGCGGTTGCTATCTCCAGTTGTGTAAGTAAAG 1129
Db 1004 ACAGTACTTCAATGCTGAGAAATGCTGCAATGATTTACGAACTTCCAGATGATAC 1063
QY 1130 GTCTCCCATAGACCAAGTAGGTGCGATGATGTTGGCTTTTGCATTTGAATGATCTGA 1189
Db 1064 GCATACCTCTCGATGATGTGACAGGAGCGGAGAGGTAAAGAGGTGAGTGGCTGCTGT 1123
QY 1190 TACGCGCGGCTTTAAGATGAAGAGCTTGTGCGGACCGCTGTAGCGCCCGAGAGAGA 1249
Db 1124 TAAGGAGAGCACAAACTTTGGCGAACTTGCACCCCAACCCAGAGAG-----ATGCGCGACA 1180
QY 1250 CTTACCGTGGCGCTATAGTTCTTTCGAGCGCTTTCGAGAGCGTSCACGAGAAATATAGCCGTAC 1309
Db 1181 GCTATGAGGAGGACTTCTGCTCGAGCCCGCAAGAGGATTTGCATGAGAGCTAATCTGCC 1240
QY 1310 TCGACTTTAGCTCGATGTACCAAAACATCATGATAAGTAAAGTAAATGTTGCTGACAGC 1369
Db 1241 TGGACTTTGCGTCCATGTTATCCCTCAATATGATTTTCATACAAACATCAGCCCGACAGC 1300
QY 1370 TTGTGAGCGCTGCTGAAAGGTGCGGAGTGTGCTGCGAGGCGCCCGAGGCTCAAGC 1429
Db 1301 TTGT-----AATAGGCAATGCGACGATTCGCAATGATGAGCGCGAGAGTGGGC 1348
QY 1430 ACAGGTTCCGTAGGTGCGCGCGCTTCTTCAAGACAGTCTTTCGAGAGGCTGTTAGAGC 1489
Db 1349 ACAAAATTCAGGAAACATCTGATGGTTTTTTTCAAAAGAAATCTCAAAATGCTGATGAGA 1408
QY 1490 TTCGTAAAGCGTGTGCGTCTGAAATGAGAGATATCTTCGAGATGCCAGATATCGAC 1549
Db 1409 AAAGAAGAGAAATAAGAGAGTTATGAAACACTTGACTACAACTCGCCAGATACAGC 1468
QY 1550 TGTGATGAAAGGCGAGAGCGCTTGAAGGTTCTTTCGAAACGCTAGTTAGCGGTACATGG 1609
Db 1469 TGCTCGATATAAGCAGCAGCAACGCTGAAAGTTCTTACAAACTCGTTTTACGTTATATG 1528
QY 1610 GTTGGAGCGCGCTAGGTGCTGATGTCAGGAGTCCGCAAGGCTGTCAGCGCTGGGTA 1669
Db 1529 GGTGGAGTCTTTCGAGATGCTGCTGCAAGAGTCCGCTGAAGCTTACAAACGCGATGGGCA 1588
QY 1670 GGCACCTCATACGCGCCCATCAACATAGCTGCTGTAACCTAGCGCTCAAGGTATCTAGC 1729
Db 1589 GACATTTATCAABACATCTGCAAGATTCGGAAGAGCTTGGATTGAACTGCTATATG 1648
QY 1730 GTGACACAGATTCGCTCTTCGTGACCTATGATCGGAGAGAGGTGGAAAAATTTCACTAAA 1789
Db 1649 GGGATACAGATAGCATCTTTGTTAAAAAAGATGATTTGAGCTTGAAGAGCTCAAAAAG 1708
QY 1790 TTATAAA-----GGAGGAGCTGGGTTTCGAAATCAAGTACAGAGA 1828
Db 1709 AAGTTAAAGAGCTCATAGGTAAACTTTCCGAGAGAGATGCCAATACAAATAGAGATAGTG 1768
QY 1829 AGGTGTACAAACGCTATTCTTTTACAGAGGCTAAGAGAGGTACGCTGGCTTCTCGAG 1888
Db 1769 AATCTACGAGACAAATATTTCTGCTGAA-----AAGAAAAAGTATGCTGGATTGACACAG 1825
QY 1889 ACAGAGGTATAGATATTGTCGGTTTCGAGGCTGTACGTCGCGATTTGCTGTAAGTCCCA 1948
Db 1826 ATGGAAGAAATTAATTTAAAGGCTCTTGAAGTTCAGAGAGGAGCGAGCTGCTGGAGCTTGA 1885
QY 1949 AGGAGGTTACAGATAAGGTTGTCGAAATAGTATTGGAAGACGAGTGGAGTGAACAGGCTG 2008

Db 1886 AGAAGATACAGAAAGGTGTAAATAGAAATCATCTCTGAAGGAAAGAAAGATCTCTGAAAGGCTG 1945
Qy 2009 TAGAGTACGTCAGAGATTTGTAAGAGTTGGAGGAGGCGAAGGTTCCCATAGAGAGC 2068
Db 1946 CTGAGTATGTGAAGAGGTATAGAGAGATGAAGGCGAAGGAAATTTCCGCTTGAAGATT 2005
Qy 2069 TTGTAATCTGGAGAGACCTTTAGTAAGCGTCTTGAGGAGTACACAAACGGAGGCCACACAG 2128
Db 2006 ATATCATCTCAAGGATTCACGAGAAACCAATCAAGTACGAGATATGAGGCTCAGC 2065
Qy 2129 TCGTTGAGGAGAGAGATCTGTACAGGCTTACCGGTAACCCAGGCGACAGATAG 2188
Db 2066 TAAAGCTGCTAGAGGCGCAAGAGAGAGAAATAGTATACAAATCGGCTCAAGGTTG 2125
Qy 2189 GGTATGTAAGTGAAGGTTGGCGGTATCATGATCAAGAGATGCGCATCTTATG 2248
Db 2126 GTTTTGTCTTCAAAAGGTTGGGAAACATAGGTATAGGCTTTTCCATCTGATCTGA 2185
Qy 2249 TCAAGGA 2255
Db 2186 TAGAGGA 2192

RESULT 10

US-10-227-110-100
; Sequence 100, Application US/10227110
; Publication No. US20030143577A1
; GENERAL INFORMATION:
; APPLICANT: Strategene
; TITLE OF INVENTION: HIGH FIDELITY DNA POLYMERASE COMPOSITIONS AND USES THEREFOR
; FILE REFERENCE: 25436/2155C
; CURRENT FILING DATE: 2002-08-23
; PRIOR APPLICATION NUMBER: 10/079,241
; PRIOR FILING DATE: 2002-02-20
; PRIOR APPLICATION NUMBER: 10/208,508
; PRIOR FILING DATE: 2002-07-30
; PRIOR APPLICATION NUMBER: 10/035,091
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 100
; LENGTH: 2325
; TYPE: DNA
; ORGANISM: Pyrococcus sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1159)..(1161)
; OTHER INFORMATION: Ns at position 1159-1161 may be any nucleotides so that the codon
; OTHER INFORMATION: at 1159-1161 encodes for Ala or Pro.
US-10-227-110-100

Query Match 7.9%; Score 190; DB 12; Length 2325;
Best Local Similarity 47.9%; Pred. No. 7.3e-51;
Matches 852; Conservative 0; Mismatches 858; Indels 69; Gaps 8;
Qy 622 GTGAAGACTGACATGGCGATGAGTGCTATTTCATTGTCAGAGGSCAAGACCATCGAAA 681
Db 508 GTGATTAAGTGAAGAGGATGCTCCCTCAGTTGAGCTGTCTCGACGGAGGGAG 567
Qy 682 CGGATACGGAGTTTGTAGAGTACGTGAAGAGATGATGACCCCGACATATAGTCGGTTAT 741
Db 568 ATGATAAAGCGTTCTCCCGTGTGTGAAGGAGAAAGACCCGAGCTTCTATACCTAC 627
Qy 742 AACAAATCAATTTGATTTGGCTTATCTTTTGAAGCGCGCCCGCATCTTAGGCATAAG 801
Db 628 AACGGGCAAACTTCGACTTCGCTTATCTGAAAAGAGCGCTGTGAAAAGCTCGAATAAC 687
Qy 802 CTTGATGTGACTAGAGAGTTGGCGCGGACCCACCATAGCGTACATGGGCACTCTCT 861
Db 688 TTCGCCCTCGAAGGATGGAAGCGGACGGAAGATTTCAGAGGATGGGCGACAGGTTTGC 747

Qy 862 GTCCCTGGCAGGCTTAACTAGATCTGTACGCTATGCGGAGAGATCCAGAGATCAAG 921
Db 748 GTGGAAGTGAAGGAGCGGATACCTTCGATCTCTATCTGTGTATGAAGACGACGATAAAC 807
Qy 922 ATAAAGAGTCTCGAGGAGTGCAGAGTATCTAGCGGTGATGAAGAGAGTGAACCGCTT 981
Db 808 CTGCCACATACACCGCTTGAAGCGCTTATGAAGCGCTTCTCGGTCACCCGAA---GGAG 864
Qy 982 ATCATCAATTTGGTGGAGATTCAGAGATTTGGGACGACCCGAGAGAGACCTATTATTA 1041
Db 865 AAGGTTTACGCTAGAGAAATAACACAGCTGGGAAACCCGCGAGAACCTTTGAGAGATC 924
Qy 1042 CTGCAATACGGCGCGAGATGTCGCGCTACTTACGCTTACGCGCTTAGCGGAGAGATTTGCG 1101
Db 925 GCCCGCTACTCGATGGAAGATCGGAGGTCAATACAGAGCTTGGGAAGGATTTCTTCCG 984
Qy 1102 TTGCTATTCAGTTGTCTGATGAACAGGTCTCCCACTAGACGAGTGTAGTGCATGAGT 1161
Db 985 ATGAGGCGCCAGCTTTCGCTTAATCGGCCAGTCCCTCTGGAGCGTCTCCGCTCCAGC 1044
Qy 1162 GTTGGCTTTGACTTGAATGTACTGATACGCGCGGCTTTTAAGATGAAGAGCTTGTG 1221
Db 1045 ACTGGCAACCTCGTTGAGTGTCTCTCAGGAAGGCTTATGAGAGATGAGCTGCGC 1104
Qy 1222 CGAAACGCGTTGAGC-----GCCCAAGAGAGACTTACCGTGGCGCTATA 1266
Db 1105 CGAAACAGCCCGATGAAGAGAGTGTGCCAAGAGACGCGAGAGCTATGAAGGANNNTAT 1164
Qy 1267 GTTCTTGAGCGCTTGAGAGGCGTGACGAGAAATATAGCGCTACTCGACTTTAGCTCGATG 1326
Db 1165 GTAAAGAGCGCGAGAGAGGCTTGTGGGAGAACATGTGTACCTAGATTTTATGATCCCTG 1224
Qy 1327 TACCCAAATCATGATGAAGTACAATGTTGGTCTCTGACAGCTTGTGAGCGCTGGTAA 1386
Db 1225 TACCCCTCAATCATCATCACCCACAACGCTCTCGCGGATACGCTCAACAGAGAGGATGC 1284
Qy 1387 AAGTGTGGCGAGTGTGTTGCTGGAGCGCCCGAGGTCAAGACAGCTTCCGTAGGTGT 1446
Db 1285 AAGGAATATGAGCTT-----GCCCAAGAGTGGCGCCACCGCTTCTGCAAGGAC 1332
Qy 1447 CGCGCGGCTTCTTCAAGACAGATTTCTGAGAGGCTGTTAGAGCTTCTGTAAGCGTGTGCT 1506
Db 1333 TTCCGAGGATTTATCCCGAGCGCTGTTGGAGACCTCTTAGAGGAGAGCGA---GATA 1389
Qy 1507 GCTGAATGAAGAGATATCTCCGAGATAGCCCAAGATATCGACTTGTGGATGAAGGAG 1566
Db 1390 AAGAAGAGATGAAGGCGCACGATTACCCGATCAGAGAGAGCTCTCGATTACAGGCGAG 1449
Qy 1567 AAGCGTTGAAGGTTCTTGCAACGCTAGTTACGCGCTACATGGTGGAGCGCGCTAGG 1626
Db 1450 AGGCGCATCAAGATCTTGGCAACAGCTACTACGCTTACTACGCTATGCAAGGCGCGC 1509
Qy 1627 TGGTATTGCGAGGAGTGGCAAGGCTGTCAAGCTTGGGCTAGCGCTCATACCGCAC 1686
Db 1510 TGGTACTGCAAGGAGTGTGCAAGAGCGTAAACGCGCTGGGAGAGGAGTACATACCGATG 1569
Qy 1687 GCGATCA---ACATAGCTGTAACTAGAGCTCAAGGTATCTACGCTGACACAGATTCG 1743
Db 1570 ACCATCAAGGAGATAGAGGAAAGTACGCGCTTTAAGGTAACTACAGCGACACCGAGGA 1629
Qy 1744 CTCTTCTGACCTATGATCCGAGAGGTGGAATTTTCATCAAAATTAAGAGGAG 1803
Db 1630 TTTTTCGCAATACCTCGAGCGGATGCTGAAACCGTCAAAAGAGGCTATGGAGTTC 1689
Qy 1804 CTGGGGT-----TCGAAATCAAGCTAGAGAGGCTGTAC 1836
Db 1690 CTCAAGTATATCAAGCCAACTTCGCGCGCGCTTGGAGCTCGAGTACGAGGCTTCTAC 1749
Qy 1837 AAAGCGTATTCTTTACAGAGGCTAAGAGAGGTACGCTGCGCTCTTCGAGGAGCGACGT 1896
Db 1750 AAACGCGCTTCTTC---GTCAAGAGAGAGATGTCGCGTGTAGACGAGGAGGAGCAAG 1806
Qy 1897 ATAGATATTGCGGTTTCGAGGCTGTACGTTGGCGATGTTGTAACCTCGCCAAAGGAGTT 1956

```

Db 1807 ATACACACCGCGGACCTGAGTTGTGAGGCGTGTGAGCGAGTAGCGAAGAGAGC 1866
QY 1957 CAGACTAAGGTTGTCGAATAGTATTGAAGACGAGTGAAGTGAACAAGGCTGTAGAGTAC 2016
Db 1867 CAGGCGAGGTTCTTGAAGCTTTGCTAAAGGACGCTGACGTCGAGAGGCGCTGAGGATA 1926
QY 2017 GTACGAGAGATTGTGAAGAGTTGAGAGGCGGAGGTCCTCCATAGAGAGCTTGAATC 2076
Db 1927 GTCAAGAGATTGTACCGAAGCTGAGCAAGTACGAGGTTCCCGGAGAGAGCTGGTGATC 1986
QY 2077 TGGAGAGCCCTTAGTAAAGCTTTGAGGAGTACAAACGAGGCGACACACGCTGTTGCA 2136
Db 1987 CACGAGCAGATAACGAGGATTAAAGGACTACAAAGGCAACCGGTCCCGCTTGCCTT 2046
QY 2137 GCGAGAGAGTGTGTACGAGGCTACCGGTAAAGCCGAGCGCAAGATAGGTTATGTA 2196
Db 2047 GCCAAGAGGTTGGCGCGAGAGGAGTCAAAATACGCGCTGGAACGGTGAATACATC 2106
QY 2197 ATAGTGAAGGTTGTGGCGCTATCAGTCAAAAGACATGGCCATATCTTCAATGTCAGGAT 2256
Db 2107 GTGCTCAAGGCTCTGGGAGGTAGGCGACAGGCGGATACGTTGAGAGGTTGACCCG 2166
QY 2257 CTTAGCCAGATAGAGTGACCC---TACTATGTTGACCAACCAATCATCCCGCTGCTGATG 2313
Db 2167 ACGAAGCACAAGTACGAGCGCGAGTACTACATTGAGAACCCAGGTTCTCCAGCGCTTGAG 2226
QY 2314 AGAATACTGGGCTACTTTGGCATCACCGAGAGAGGCTG 2352
Db 2227 AGAATCTGAGAGCCTTCGTTACCGCAAGGAGACCTG 2265

```

RESULT 11

```

US-10-208-508-100
; Sequence 100, Application US/10208508
; Publication No. US20030180741a1
; GENERAL INFORMATION:
; APPLICANT: Stratagene
; TITLE OF INVENTION: HIGH FIDELITY DNA POLYMERASE COMPOSITIONS AND USES THEREFOR
; FILE REFERENCE: 25436/2155B
; CURRENT APPLICATION NUMBER: US/10/208,508
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: 10/079,241
; PRIOR FILING DATE: 2002-02-20
; PRIOR APPLICATION NUMBER: 10/035,091
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 100
; LENGTH: 2325
; TYPE: DNA
; ORGANISM: Pyrococcus sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1159)..(1161)
; OTHER INFORMATION: Ns at position 1159-1161 may be any nucleotides so that the codon
; OTHER INFORMATION: at 1159-1161 encodes for Ala or Pro.
US-10-208-508-100

```

```

Query Match      7.9%; Score 190; DB 12; Length 2325;
Best Local Similarity 47.9%; Pred. NO. 7.3e-51;
Matches 852; Conservative 0; Mismatches 858; Indels 69; Gaps 8;

QY 622 GTGAGACTGACGATGGGATGAGGTGCTATTTCATTGCGAGGCGCAAGACGATCGAAA 681
Db 508 GTGATACTTGGAGAGCTGTGATCTCCCTACGTTGACGTCTCGACGAGAGGGAG 567
QY 682 CGGATACCGGAGTTGTAGAGTACGTGAAGAGGTATGACCCCGACATAATAGTGGTTAT 741
Db 568 ATGATAAAGCGCTTCTCCGTGTTGTGAAGAGAGAAAGACCGGACGTTCTCATACCTAC 627
QY 742 AACACATCATCTCGATGGGCTTATCTTTTGGGCGCGCCGCCATCTTAGCATPAAG 801

```

```

Db 628 AACCGGACAACTTCGACTTCGCTATCTGAAAAAGCGCTGTGAAAAGCTCGAATAAAC 587
QY 802 CTTGATGTACTAGAGAGTTGGCGGAGCCACCACACTAGCGTACATGGGCACGTCCT 861
Db 688 TTCCCTCTCGAGAGGATGGAAGCGAGCCGAGANTTCAGAGGATCGGCACAGGTTGGC 747
QY 862 GTCCCTGCGAGCTTAACGTAGATCTGTACGACTATGTCGCGAAGAGATCCAGAGATCAAG 921
Db 748 GTGAGTGAAGGACCGATACACTTCGATCTCTATCTGTGATAAGACGACGATTAAC 807
QY 922 ATAAAGAGTCTCGAGGAGTTCGACGAGTATCTAGGCGTGTATGAGAGAGTGAACGCGTT 981
Db 808 CTGCCACATACACGCTTGAGGCGTTTATGAAGCGGCTCTTCGCTCAGCGAA---GGAG 864
QY 982 ATCATCAATTGCTGGGAGATTCCAGACTATTGGGACGACCCGAGAGAGACACTATTTA 1041
Db 865 AAGGTTTACGCTGAGGAAATAACACAGCTTGGGAAACCGCGCAGAGAACTTGAAGAGTTC 924
QY 1042 CTGCAATACGCGCGACGATGTCGCGCTACTTACGCGTTAGCCGAGAGATATTGCCG 1101
Db 925 GCGCGTACTCGATGGAAGATGGAAGTCCACATACGAGCTTGGGAGGAGTTCTTCCG 984
QY 1102 TTTGCTATCCAGTTGTCGTACGTAAACAGGTCCTCCACTAGACACAGTGTGGGATGAT 1161
Db 985 ATGAGGCGCCAGCTTCTCGCTTAATCGGCGAGTCCCTCTGGGACCTCTCCCGCTCCAG 1044
QY 1162 GTTGGCTTTCGACTTGAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 1221
Db 1045 ACTGGCAACCTCGTTGAGTGTCTCTCTCAGAGAGGCTTATGAGAGGATGAGCTGGCC 1104
QY 1222 CCGAACCCTGTTGAGC-----GCCGAGAGAGACTTACCGTGGCGCTATA 1266
Db 1105 CCGAACCCTGTTGAGC-----GCCGAGAGAGACTTACCGTGGCGCTATA 1266
QY 1267 GTTCTTGAGCGGTTGAGAGCGCTGACGAGATATATATATATATATATATATATATAT 1326
Db 1165 GTAAAGAGCGCGAGAGGTTGTGGGAGAAATATATATATATATATATATATATATAT 1224
QY 1327 TACCCAAACATCATGATAAAGTACAATGTTGGTCTTGACACGCTTGTGAGGCGCTGTA 1386
Db 1225 TACCCCTCAATCATCATCAACCCACAGCTCTCGCGGATACGCTCAACAGAGAGGATGC 1284
QY 1387 AAGTGTGGGAGTGTGTTGCTGGAGGCGCGGAGGTCGACGACAGGTCCTAGAGTGT 1446
Db 1285 AAGGAATATGAGCTT-----GCCGAGAGGCTCGGCGCCGCTTCTGCAAGAG 1332
QY 1447 CCGCCCGCTTCTTCAAGACAGTCTTGTGAGAGGCTGTGTAGAGCTTGTGTAGCGTGTGCGT 1506
Db 1333 TTCCGAGGATTTATCCGAGCGCTTGTGAGAGCTCTCTGAGAGAGGCGAGAA---GATA 1389
QY 1507 GCTGAATGAAGAGATATCTCCGAGTAGCCGAGATATCGACTGTGATGAAGAGGAG 1566
Db 1390 AAGAGAGAGATGAAGAGGACGATGACCCGATGAGAGAGGAGCTCTCGATTAAGCGAG 1449
QY 1567 AAGCGTTTGAAGGTTCTTCAAGACGCTAGTTAGCGGTACATGAGGTTGAGCGCGCTAG 1626
Db 1450 AGGGCCATCAAGATCTTGGCAACAGCTACTACGCTTACTACGCTATGCAAGGCGCGC 1509
QY 1627 TGGTATTGCGAGGAGTTCGCAAGGCTGTCTACGCGCTTGGGTTAGGCACTTACACGACC 1686
Db 1510 TGGTACTGCAAGGAGTGTGAGAGAGGCTTAACGCGCTTGGGAAAGGAGGATACATAAG 1569
QY 1687 GCCATCA---ACATAGCTCGTAAACTAGGCTCAAGGTGATCTACGCTGACACAGATTCTG 1743
Db 1570 ACCATCAAGGAGATGAGAGGAAAGTACGCTTTTAAAGTATCTACGCGACACCGAGGA 1629
QY 1744 CTCTTCGTGACCTATGATCCGAGAGAGGTCGAAATTTTCATCAAAATTTTAAAGGAGGAG 1803
Db 1630 TTTTTCGCAACATACCTGGAGCGGATGCTGAAACCGTCAAAAGAGGCTATGGAGTTC 1689
QY 1804 CTGGGCT-----TCGAAATCAAGCTAGAGAGGTTGATC 1836
Db 1690 CTCAAGTATATCAACGCGCAAACTTTCGCGGCGCGCTTGTAGCTGAGTACGAGGCGCTTCTAC 1749

```

QY 1837 AAACGCTTATTTTACAGAGCTAAGAGAGCTAGCTGGCTTCTCGAGAGCGAGCT 1896
Db |||||
1750 AAACCGGGCTTCTTC---GTCAAGAGAAAGATATGCGGTATAGACGAGAGCGAAG 1806
QY 1897 ATAGATATTTGTCGGTTTCAGAGCTGTACGTGGCGATTGGTGTGAACCTGCGCAAGGAGTT 1956
Db |||||
1807 ATACACACCGGGAGCTTGGATTGTGAGCGTGTACTGGAGCGAGATAGCGAAGAGAG 1866
QY 1957 CAGACTAAGGTTGTGGAATATGTTGAAGACGAGTGGTGAACAAGGCTGTAGAGTAC 2016
Db |||||
1867 CAGGCGAGGTTCTTGAAGCTTTGCTAAAGGACGCTGACGTCGAGAGAGCGCGTGAAGATA 1926
QY 2017 GTCAAGAGATTGTGAAGAGTGTGAGGAGGCGCAAGGTTCCCATAGAGAGCTTGTATC 2076
Db |||||
1927 GTCAAGAGATTGTGAAGAGTGTGAGGAGGCGCAAGGTTCCCGCGAGAGAGCTGTGTATC 1986
QY 2077 TGAAGACGCTTAGTAAAGGCTTGTAGGAGTACACAAAGGAGGACACAGCTGCTTGA 2136
Db |||||
1987 CAGGACGATACAGAGGATTTAAAGGACTACAGGCAACCGTCCCAAGCTTCCGTT 2046
QY 2137 GCGAAGAGATGCTGTACAGGCTACCGGGTAAGCCGAGCGACAGATAGGTAATGA 2196
Db |||||
2047 GCGAAGAGTGTGCGCGAGAGGAGTCAAAATACCGCTTGAAGCGTGTATAGCTACATC 2106
QY 2197 ATAGTAGAGGTTGGTGGCGTATCAGTCAAGAGCATGCGCATCTTATGCTCAAGAT 2256
Db |||||
2107 GTGCTCAAGGCTCTGGGAGGATAGGCGACAGGCGGATACCGTTCGACGAGTTCAACCG 2166
QY 2257 CCTAGCCAGATAGAGTGAC---TACTATGTTGACCAACAAATCATCCCGCTGCATTG 2313
Db |||||
2167 AGAAGCACAAGTACAGCGCGAGTACTACATTGAGAACAGGTTCTCCAGCGCTTGA 2226
QY 2314 AGAATCTGGCTACTTTGGCATCACCGAGAGAGAGCTG 2352
Db |||||
2227 AGAATCTGAGAGCTTGGTTTACCGCAAGGAGAGCTG 2265

RESULT 12
US-09-852-922-3
; Sequence 3, Application US/09852922
; Patent No. US20020076768A1
; GENERAL INFORMATION:
; APPLICANT: TOYO BOSEKI KASUHIKI KAISHA
; TITLE OF INVENTION: MODIFIED THERMOSTABLE DNA POLYMERASE
; FILE REFERENCE: 000053
; CURRENT APPLICATION NUMBER: US/09/852,922
; CURRENT FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 2000-138796
; PRIOR FILING DATE: 2000-05-11
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2325
; TYPE: DNA
; ORGANISM: Pyrococcus kodakaraensis
US-09-852-922-3

Query Match 7.9%; Score 189.8; DB 9; Length 2325;
Best Local Similarity 47.9%; Pred. No. 8.5e-51;
Matches 853; Conservative 0; Mismatches 857; Indels 69; Gaps 8;

QY 622 GTGAGACTGACGATGGGATGAGTGCTATTTCATTGTCAGAGGCGCAAGAGCATCGAAAA 681
Db |||||
508 GTGATAACTTGGAGAACGTTGGATCTCCCTACGTTGACGTCGTCGACGAGAGGAG 567
QY 682 CGATACCGGAGTTGTAGTACGTTGAGAGGATGATGCCCGCATATATAGTCGGTAT 741
Db |||||
568 ATGATAAAGCGCTTCTCCGCTGTGTGAGGAGAGAAAGCCCGAGCGTTCTCATACCTAC 627
QY 742 AACCAACATCAATTTTCGATTGGCTTATCTTTTGGCGCGCGCGCATCTTAGGATAAAG 801
Db |||||
628 AACGCGCAACTTGCCTTCTGCTTATCTGAAAGAGCGCTGTGAAAGAGCTCGGATTAAC 687

QY 802 CTTGATGTGACTAAGAGATTGGCGCGAGCCACCACTAGCGTACATGGSCACGCTCTCT 861
Db |||||
688 TTGCGCTTCGAGAGGATGGAAGCGAGCGAAGATTACAGAGGATGGCGACAGCTTTGCC 747
QY 862 GTCCCTGGCAGGCTTACGCTAGATCTGTACACTATCCGAGAGATGCCAGAGATCAAG 921
Db |||||
748 GTCGAAGTGAAGGACCGATACACTTCGATCTCTATCTGTATTAACGAGCGATTAAC 807
QY 922 ATAAAGAGTCTCGAGGAGGTCGACAGATATCTAGGCTGTATGAAGAGATGACCGCTT 981
Db |||||
808 CTGCCCCACATACAGCTTGTAGGCGCTTTATGAAGCGCTCTTCGCTAGCCGAA---GGAG 864
QY 982 ATCATCAATTTGGTGGAGATTCCAGACTATTGGGACGACCCGAGAGAGACCACTATTA 1041
Db |||||
865 AAGTTTTACGCTGAGGAAATAACACACAGCTGGGAAACCGCGAGACCTTTGAGAGATC 924
QY 1042 CTGCAATACCGCGCGAGGATGTCGCGCTACTTACCGCTTAGCCGAGAGATATTGCCG 1101
Db |||||
925 GCCCGCTACTCGATGGAGATGCAAGGTCACTATACGAGCTTGGGAGGAGTTCCTTCG 984
QY 1102 TTTCTATCCAGTTTCTGTACGTAAACAGGTCCTCCACTAGACAGGTTAGGTGCGATGAT 1161
Db |||||
985 ATGAGGCGCCAGCTTCTCGCTTAATCGCGAGTCCCTCTGGGACGCTCTCCCGTCCAGC 1044
QY 1162 GTTGGCTTTTCGACTTGAATGTTACTGATACGCGCGCTTTAAGATGAAGAGCTTTGTG 1221
Db |||||
1045 ACTGGCAACCTCGTTGAGTGGTTCTCTCCTCAGAAAGGCTTATGAGAGGATGAGCTGGCC 1104
QY 1222 CCGAACCGCTTGGC-----GCCAGAGAGACTTTACCGTGGCGCTATA 1266
Db |||||
1105 CCGAACAGCGCCGATGAAAAGGAGCTGGCCAGAGAGCGGACAGCTATGAAGGAGGCTAT 1164
QY 1267 GTTCTTCCAGCTTGTAGAGCGCTGACAGAAATATACCGCTACTCCGACTTTTACGTCATG 1326
Db |||||
1165 GTAAAGAGCGCGAGAGAGGTTGTGGGAGAACATAGTGTACCTAGATTTTATGATCCCTG 1224
QY 1327 TACCAAAACATCATGATAAAGTACAAATGTTGGTCTGTACACGCTTGTAGCGCTGTGAA 1386
Db |||||
1225 TACCCCTCAATCATCATCACCCACACGCTCGCCGAGATACGCTCAACAGAGAGGATGC 1284
QY 1387 AAGTGTGGCAGTGTGTTCTGGAGGCGCCCGAGAGTCAAGCACAGGTTCCGATGCTGT 1446
Db |||||
1285 AAGGAATATACGCTT-----GCCCAACAGTCTCGCCACCGCTTCTCAAGAGAC 1332
QY 1447 CCGCCCGCTTCTTCAAGACAGTCTTGTAGAGGCTGTAGAGCTTCTGTAAAGCGTGTGCGT 1506
Db |||||
1333 TTCCAGGATTTATCCGAGCTTGTGGAGACTCTTAGAGGAGAGGAGAA---GATA 1389
QY 1507 GCTGAATGAAGAGTATCTCCGATAGCCAGATATCGACTTGTGGATGAAGGCGAG 1566
Db |||||
1390 AAGAAGAGATGAAGGCGACGATTACCCGATCGAGAGGAGCTCTCGATTACAGGCGAG 1449
QY 1567 AAGCGTTGAGGCTTCTGCAAGCGTAGTTACGGCTACATGGGTTGGAGCGCGCTAGG 1626
Db |||||
1450 AGGCGCATCAAGATCTCGCAAAACAGCTACTACGCTTACTACGCTTATGCAAGGCGCGC 1509
QY 1627 TGTATTGTCAGGAGTGGCAAGGCTGTCCAGCGCTTGGGTAGGCGACCTCATACGACC 1686
Db |||||
1510 TGTACTGTCAAGAGGTTGTCAGAGCGCTACGCTGCGAAGGAGGATCATTAAGCATG 1569
QY 1687 GCATCA---ACATAGCTGTAACTAGGCTTCAAGGCTTACCGTATGACACAGATTCG 1743
Db |||||
1570 ACCATCAAGGAGATAGAGGAAAGTACGCTTTTAAAGTAACTTACAGGACACCGAGGA 1629
QY 1744 CTCTTGTGACCTTATGATCGGAGAGGTTGGAATTTTCATCAAAATTTATAAGAGGAG 1803
Db |||||
1630 TTTTGTGCCAATAATCTGAGCGCGATGCTGAAACCGTCAAAAGAGGCTATGAGGTTTC 1689
QY 1804 CTGGGT-----TCGAAATCAAGCTAGGAGGTTGATAC 1836
Db |||||
1690 CTCAAGTATATCAACGCCAAACTTCGCGGCGCTTGTAGCTGAGTACGAGGCTTCTTAC 1749

Db 1690 CTCAAGTATATCAACGCCAACTTCGGCGCGCTTGAGCTCGATGAGGGCTTCTAC 1749
Qy 1837 AAACGCTTATCTTTAAGAGGCTTAAGAGGCTACGCTGCGCTTCTCGAGGAGGAGCT 1896
Db 1750 AAACGCGCTTCTTC--GTACACGAAGAAGATATCGGTGTATAGACGAGGAGGCAAG 1806
Qy 1997 ATAGATATTGTCGGTTCGAGGCTGTACGTGGCGATTGGTGAACCTCGCAGAGGCTT 1956
Db 1807 ATAAACAGCGCGGACTTGTGANTGTGAGCGTGTGAGCGAGATAGCAAGAGAGAG 1866
Qy 1957 CAGACTAAGTGTTCGAATAGTATTGAAGACGAGTGAAGGCTGAACAGGCTGTAGATAC 2016
Db 1867 CAGCGAGGCTTCTTGAAGCTTCTTAAGACGCTGACGTCGAGAGGCGGTGAGGATA 1926
Qy 2017 GTCAGGAGATGTGAAGAGTTGGAGGAGGCAAGGTTCCCATAGAGAGGCTGTAAATC 2076
Db 1927 GTCAAGAAGTATACCGAAAGCTGACCAAGTACGAGGTTCCGCGGAGAGAGCTGTGATC 1986
Qy 2077 TGGAGACCCCTAGTAAAGCGTCTGAGGCTGACACAGGAGGAGGAGGAGGAGGAGG 2136
Db 1987 CACGAGCAGATACGAGGAGTTTAAGAGCTTACAGGCAACCGGTCCCGCTTCCGCTT 2046
Qy 2137 GCGAAGAGGATGTCTGACAGGCTACCGGTAAAGCGGACGAGGAGGAGGAGGAGGTA 2196
Db 2047 GCCAAGAGGTTGGCGCGAGAGGAGTCAARAATACGCTTGGAAACGGTGATAAGCTATC 2106
Qy 2197 ATAGTGAAGGTTGGTGGCGGTATCAGTCAAGAGCATGCGCATCTTCATGTTGTCAGGAT 2256
Db 2107 GTGCTCAAGGCTCTGGAGGATAGGCGACAGGCGGATACCGTTTCGACGAGTTTCGACCG 2166
Qy 2257 CCTAGCCAGATAGACGTGACG--TACTATGTTGACCAACCAATCATCCCGCTGCAATTG 2313
Db 2167 ACGAAGCAGATGACGCGGAGTACTACATNTGAGAACAGGTTCTCCAGCGCTTGAG 2226
Qy 2314 AGAATCTGGCTACTTTGGCATCACCGAGAGAGCTG 2352
Db 2227 AGAATCTGAGACCTTCGGTTACCGCAAGGAGACCTG 2265

RESULT 14

US-10-208-508-97
; Sequence 97, Application US/10208508
; Publication No. US20030180741A1
; GENERAL INFORMATION:
; APPLICANT: Stratagene
; TITLE OF INVENTION: HIGH FIDELITY DNA POLYMERASE COMPOSITIONS AND USES THEREFOR
; FILE REFERENCE: 25436/2155B
; CURRENT APPLICATION NUMBER: US/10/208,508
; CURRENT FILING DATE: 2003-02-07
; PRIOR FILING DATE: 2003-02-07
; PRIOR FILING DATE: 2002-02-20
; PRIOR APPLICATION NUMBER: 10/079,241
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 2325
; TYPE: DNA
; ORGANISM: Pyrococcus sp.
US-10-208-508-97

Query Match 7.9%; Score 189.8; DB 12; Length 2325;
Best Local Similarity 47.9%; Pred. No. 8.5e-51;
Matches 853; Conservative 0; Mismatches 857; Indels 69; Gaps 8;
Qy 622 GTGAGACTGACGATGGCGATGAGTGTATTCATTTGAGAGGCGCAAGAGCATCGAAAA 681
Db 508 GTGATACTTGGAGAGCGTGTGATCTCCCTACGTTGACGTCGTCTCGACGAGAGGAG 567
Qy 682 CCGATACCGGAGTTGTAGATGACGTGAAGAGGATGACCCCGACATAATAGTCGGTTAT 741
Db 568 ATGATAAGCGCTCTCCTCGGTGTGAGGAGAGAGACCGGACGTTCTCATAAACCTAC 627

Qy 742 AACAAATCATTTGCTTGGCTTATCTTTTGGCGCGCGCGCATCTTAGCATTAAG 801
Db 628 AACGGGCAACTTTCGACTTCGCTTATCTGAAAAAGCGCTGTGAAAGCTCGGAATAAC 687
Qy 802 CTGTGATGTACTAGAGAGTTGGCGCGAGCCACCACTAGCGTATACATGGGCACGCTCT 861
Db 688 TTGCGCTCGGAAGGATGGAAGCGAGCGAGATTCAGAGGATGGCGACAGGTTTGGC 747
Qy 862 GTCCCTGGGAGGCTTAAAGTAGATCTGTACGATATGCGGAGATGCGGAGAGATCCAGATCAAG 921
Db 748 GTCGAAGTGAAGGAGCGGATACACTTCGATCTCTATCTGTGTATGAAGACGAGCATAAAC 807
Qy 922 ATAAAGAGTCTCGAGGAGGCTCGACAGTATCTAGGCGGTATGATGAAGAGTGAACGCGTT 981
Db 808 CTGCCCCACATACACGCTTGAGGCGGTTTATGAAGCGGCTTTCGTCAGCCGAA---GGAG 864
Qy 982 ATCATCAATGTGGTGAGATTCAGACTATTGGGACGACCCGAAAGAGAGCACTATTATTA 1041
Db 865 AAGGTTTACGCTGAGGAATTAACACAGCGCTGGGAAACCGGCGAGAACTTTGAGAGATC 924
Qy 1042 CTGCAATACGCGGGGAGCATGTTCGCGCTACTTACGCGCTTAGCGGAGAGATATTGCGC 1101
Db 925 GCCCGCTACTCGATGGAAGATCGAAGGTTCATACGAGCTTGGGAAGGAGTTCTCTCG 984
Qy 1102 TTGTCTATCCAGTTGTGCTACGTTAAACAGGTCTCCCACTAGACCAAGTGGTGGATGAT 1161
Db 985 ATGGAGGCCAGGCTTTCGCTTATTCGCGCCAGTCCCTCTGGGAGCTTCCCGCTCCAGC 1044
Qy 1162 GTTGGCTTTCGATTTGAATGTGATACCTGATACGCGGCGCTTTTAAGATGAAGAGCTTGTG 1221
Db 1045 ACTGGCAACCTCGTTGAGTGGTTCTCTCTCAGGAAGGCTATGAGAGGAATGAGCTGGCC 1104
Qy 1222 CGAAGCCGCTTGAGC-----GCCGAGAAAGACTTACCGTGGCGCTATA 1266
Db 1105 CGAACAAGCCCGATGAAAGGAGCTGGCCAGAGACGCGCAGAGCTATGAAGAGAGCTAT 1164
Qy 1267 GTTCTTTGAGCGTTGAGAGCGGTGACGAGAGATATAGCCGTACTCGACTTTTACGTCGATG 1326
Db 1165 GTAAAGAGCCGAGAGAGGTTTGGGAGAAACATAGTGTACTAGATTTTATAGTCCCTG 1224
Qy 1327 TACCCAAACATCATATAAGTACATTTGCTGCTGACACGCTTGTGAGGCTTGTGAA 1386
Db 1225 TACCCCTCATCATCATCAACCCACACGCTCTCCCGGATACGCTCAGAGAGAGATCG 1284
Qy 1387 AAGTGTGGGAGTGTGTTGCTGGGAGGCGCGGAGGCTCAAGACAGGTTCCGTAGGTGT 1446
Db 1285 AAGGAATATGACGTT-----GCCCCACAGGTCGCGCACCGCTTCTGCAAGGAC 1332
Qy 1447 CGCCCGGCTTCTTCAAGACAGTTCTTGAAGGCTGTGAGAGCTTCTGTAAGCGTGTGCGT 1506
Db 1333 TTCCAGAGATTTATCCGAGCGCTGTTGAGAGCTTCTTAGAGAGGAGGAGGAGAA---GATA 1389
Qy 1507 GCTGAAATGAAGAGTATCTCCGATAGCCAGAGATATCGACTGTTGATGAAGAGGAG 1566
Db 1390 AAGAAGAAGTGAAGGCGACGATTGACCGGATCGAGAGGAGCTCCCTCGATTACAGGCGAG 1449
Qy 1567 AAGCGTTGAAGTTCCTGCAAAACGCTAGTTAGCGCTACATGGTGTGAGGCGCGGTAGG 1626
Db 1450 AAGGCGCATCAAGATCTCTGGCAAAACAGCTACTACGGTTACTACGGCTATGCAAGGCGCGC 1509
Qy- 1627 TGTATTTCAGGAGGTGCGCAAGGCTGTCAGGCTTGGGAGGAGGAGGAGGAGGAGGAGG 1686
Db 1510 TGTACTGAGAGGTGTCAGAGAGCGTAAAGGCTGCGGAGGAGGAGGAGTACATTAAGCATG 1569
Qy 1687 GCATCA---ACATAGCTCTAAACCTAGGCTCAAGGCTGATCTACGGTGACACAGATTGCG 1743
Db 1570 ACCATCAAGGAGATAGAGGAAAGTACGCTTTAAGGTAATCTACAGCGACACCGACGGA 1629
Qy 1744 CTCTTCGTGACCTATGATCCGAGAGAGGTGGAAATTTTCATCAAAATTTAAGAGGAG 1803
Db 1630 TTTTTCGCCAATACCTCTGGAGCGGATGCTGAAACCGTCAAAAGAGGAGCTATGAGATTTC 1689


```
QY 1804 CTGGGGT-----TCGAAATCAAGCTAGAGAGGTGTAC 1836
Db 1690 CTCAGTATATCAAGCCAAACTCCGGCGCGCTTGAGCTCGATGAGAGGCTTCTAC 1749
QY 1837 AAACGCTTATCTTTACAGAGGCTAAGAGAGGTACGCTGGCCTCTTCGAGGACGACGT 1896
Db 1750 AAACCGGCTTCTTC---GTCAGAGAGAGAGTATGCGGTATAGACGAGAGGCAAG 1806
QY 1897 ATAGATATGTCCGTTTCAGGCTGTACGTGGCGATTGGTGTGAACCTCGCCAAAGAGGTT 1956
Db 1807 ATAAACAACGCGGAGCTTGAGATTGTGAGCGGTGACTGAGCGGAGATAGCGAAAGAGACG 1866
QY 1957 CAGACTAGGTTGTCGTAATAGTATTAGACAGGTAGGTGAACAAGGCTGTAGAGTAC 2016
Db 1867 CAGGCGAGGTTCTTGAAGCTTTGCTAAAGGACGCTGACGTCGAGAGGCGGTGAGGATA 1926
QY 2017 GTCAGAGAGATTGTGAAGAGTTGAGAGGAGGCAAGGTTCCCATAGAGAGCTTCTAATC 2076
Db 1927 GTCGAAGAAGTTACGGAAGAGCTGACCAAGTACGAGGTTCCCGCGAGAGAGCTGTGATC 1986
QY 2077 TGGAGAACCTTTAGTAAGCGTCTGAGGAGTACACAACGAGGACCAACGCTCGTTGCA 2136
Db 1987 CACGAGCAGATACGAGGATTTAAAGGACTACAAGGCAACCGGTCCCGACGTTGCCGTT 2046
QY 2137 GCGAGAGAGATCTGTACGAGGCTACCGGTTAAGCCAGCGCAAGATAGGATGTA 2196
Db 2047 GCCAAGAGGTTGCGCGAGAGAGGTCAAAATACGCCCTGGAAACGCTGATAAGCTACATC 2106
QY 2197 ATAGTGAAGGTTGGCGGTATCAGTCAAGAGCATGGCCATCTTCATGCTCAAGAT 2256
Db 2107 GTGCTCAGGGCTCTGGGAGTATGCGACAGGCGATACCGTTGACAGATTGCAACCG 2166
QY 2257 CTTACCCAGATAGAGTGACC---TACTATGTTGACCAACCAATCATCCCGCTGCAATG 2313
Db 2167 ACCAAGCACAAGTACGAGCGCGAGTACTACATTGAGAACCAAGGTTCTCCAGCGGTTGAG 2226
QY 2314 AGAATACTGGCTACTTTGGCATCACCGAGAGAGCTG 2352
Db 2227 AGAATCTGAGAGCCTTCGGTTACCGCAAGGAAGACCTG 2265
```

RESULT 15

```
US-10-227-110-98
; Sequence 98, Application US/10227110
; Publication No. US20030143577A1
; GENERAL INFORMATION:
; APPLICANT: Stratagene
; TITLE OF INVENTION: HIGH FIDELITY DNA POLYMERASE COMPOSITIONS AND USES THEREFOR
; FILE REFERENCE: 25436/2155C
; CURRENT APPLICATION NUMBER: US/10/227,110
; CURRENT FILING DATE: 2002-08-23
; PRIOR APPLICATION NUMBER: 10/079,241
; PRIOR FILING DATE: 2002-02-20
; PRIOR APPLICATION NUMBER: 10/208,508
; PRIOR FILING DATE: 2002-07-30
; PRIOR APPLICATION NUMBER: 10/035,091
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 98
; LENGTH: 2325
; TYPE: DNA
; ORGANISM: Pyrococcus sp.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1150)..(1152)
; OTHER INFORMATION: Ns at positions 1150-1152 may be any nucleotides so that the code
; OTHER INFORMATION: n at 1150-1152 encodes for Asn, Leu, His, Gln, or Ser.
US-10-227-110-98
```

Query Match

```
Best Local Similarity 7.8%; Score 188.4; DB 12; Length 2325;
Matches 851; Conservative 0; Mismatches 859; Indels 69; Gaps 8;
```

```
QY 622 GTGAAGACTGACGATGCGATGAGGTGCTATTTCATTGACAGAGGCGCAAGACGATCGAAA 691
Db 508 GTGATACTTTGGAAGAACGTTGATCTCCCTCTAGCTTGCAGTCTCTCGAGGAGGGAG 567
QY 682 CCGATACGCGAGTTTGTAGAGTACGTGAAGAGGTATGACCCCGACATAATAGTCGGTTAT 741
Db 568 ATGATAAAGCGCTTCTCTCCGTGTGTGAAGGAGAAAGACCCGACGCTCTCATAAAGCTAC 627
QY 742 AACAAACATCATTTCCGATTCGCTTATCTTTTTCGAGGCGCGCCCGCATCTCTAGGCATAAG 801
Db 628 AACGCGACATCTTCGATTCGCTTATCTGAAAAGCGCTGTGAAAAGCTTCGGAATAAAC 697
QY 802 CTTGATGTGACTAGAAGAGTTGGCGCGAGCCCAACCACTAGCCGTATCATGGGCCAGCTCTCT 861
Db 688 TTGCGCTTCGGAAGGATGGAAGCGAGCGCAAGATTACAGAGGATGGCGACAGGTTTGCC 747
QY 862 GTCCCTGGCAGGCTTAACGTAGATCTGTAGACTATGTCGCCGAGAGATGCCAGAGATCAAG 921
Db 748 GTCGAAGTGAAGGACCGATACACACTTCGATCTCTATCTCTGTGATAAGACGACGATAAAC 807
QY 922 ATAAAGAGTCTCGAGGAGGTGCGAGAGTATCTAGGCGTGTATGAAGAGTGAAGAGTGAACCGGTT 981
Db 808 CTGCCACATACACGCTTGAGGCGGTTTATGAAGCGCTCTTCGTCAGCGCAA---GGAG 864
QY 982 ATCATCAATTTGGTGGAGATTTCAGACTATTGGGACGACCCGAGAGAGAGACCACTATTA 1041
Db 865 AAGGTTTACGCTGAGGAAATAACACACAGCTGGGAAACCGCGGAGAACCTTGAGAGAGTC 924
QY 1042 CTGCATACGCGCGACGATGTCGGGCTACTTACGGCTTACCGGTTACCGAGAGATATTGCCG 1101
Db 925 GCGCGCTACTGATGGAAGATGCGAAGGTACATACGAGCTTGGGAGAGGATTCCTTCG 984
QY 1102 TTTGCTATCCAGTTGCTGATGTAACAGAGTCTCCCACTAGACAGAGTGTGGTGGATGAGT 1161
Db 985 ATGAGGCGCAGCTTCTCGTTAATCGGCGAGTCCCTCTGAGACCTCTCCCGCTCCAGC 1044
QY 1162 GTTGGCTTTCGACTTGTGATGTTACCTGATACGCGCGCGGTTTAAAGATGAAAGAGCTTGTG 1221
Db 1045 ACTGGCAACCTCGTTGAGTGTCTCTCTCAGGAGGCGCTATGAGAGGATGAGCTGGCC 1104
QY 1222 CCGAACCGCTTGAGC-----GCCAGAGAGACTTACCGTGGCGGTATA 1266
Db 1105 CCGAACAAAGCCGATGAAAAGGAGCTGGCCAGAGAGCGGACAGCAGCAGGAGGCTAT 1164
QY 1267 GTTCTTGAGCGCTTGAGAGCGGTGACGAGAAATAGCCGTACTCCGCTTTTACCTGATG 1326
Db 1165 GTAAAGAGCGCGAGAGAGGTTGTGGAGAACATAGTGTACTAGATTTTATCCCTG 1224
QY 1327 TACCCAAACATCATGATAAAGTACAAATGTTGTCCTGACACGCTTGTGAGGCTTGTGAA 1386
Db 1225 TACCCCTCAATCATCATCAACCAAGCTCTCCCGGATACGCTCAACAGAGAGGATGC 1284
QY 1387 AAGTGTGGCAGTGTGTTCTCTGGGAGGCGCGGAGGTCAAGCACAGGTTCCGTAGGTGT 1446
Db 1285 AAGGAATATGACGTT-----GCCCAAGGTCGGCCACCGCTTCTGCAAGGAC 1332
QY 1447 CCGCCCGGCTTCTTCAAGACAGTCTTGTAGAGCTGTGTAGAGCTTCTGTAAGCGTGTGCGT 1506
Db 1333 TTCCAGGATTTATCCGAGCCTGCTTGAGACCTCTGTAGAGAGGAGGAGAA---GATA 1389
QY 1507 GCTGAATGAAGAGTATCCTCCGATAGCCCAAGAAATATCGACTGTTGATGAAGGCGAG 1566
Db 1390 AAGAAGAGATGAAGCCACGATTACCCGATCGAGAGGAGGAGCTCTCCGTTTACAGGAG 1449
QY 1567 AAGCGCTTGAAGTCTTCTGAAACGCTAGTACGCTACATGCGTTGGAGGCGCGCTAGG 1626
Db 1450 AGGCGCATCAAGATCTCTGCAACAGCTACTACTGCTTACTACGGCTATGCAAGGCGCGC 1509
QY 1627 TGGTATTGAGGAGGTGCGCAAGAGGCTGTACCGCTTGGGTAGGACCTCTCATACGACC 1686
Db 1510 TGGTACTGCAAGAGGTGTGCAAGAGCGGTAAACGCGCTGGGAGAGGAGGTACATTAACGATG 1569
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2003, 12:02:29 ; Search time 141 Seconds
(without alignments)
7550.469 Million cell updates/sec

Title: US-10-034-849-1
Perfect score: 2412
Sequence: 1 atgactgaagtgtattccac.....tagccaagaagcaagtaa 2412

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues
Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCRS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2412	100.0	2412	2	US-08-907-166-3
2	2412	100.0	2412	4	US-09-391-340-3
3	1069.6	44.3	2430	1	US-08-062-368-3
4	1045.6	43.3	2430	1	US-08-062-368-1
5	279	11.6	2367	2	US-08-907-166-5
6	279	11.6	2367	4	US-09-391-340-5
7	131.4	7.9	2325	3	US-08-902-632-3
8	162.2	6.7	2337	2	US-08-906-925-5
9	146.2	6.1	2734	1	US-08-375-134-13
10	146.2	6.1	2734	5	PCT-US95-15263-13
11	131.4	5.4	5339	3	US-09-073-354-6
12	131.4	5.4	5339	3	US-08-656-005A-6
13	131.4	5.4	5339	3	US-09-073-259-6
14	131.4	5.4	5339	3	US-09-363-095-6
15	131.4	5.4	5342	3	US-08-418-027-6
16	131.4	5.4	5342	3	US-08-902-632-1
17	131.4	5.4	5342	3	US-09-073-354-5
18	131.4	5.4	5342	3	US-08-656-005A-5
19	131.4	5.4	5342	3	US-09-073-259-5
20	131.4	5.4	5342	3	US-09-363-095-5
21	131.4	5.4	5342	3	US-09-418-027-5
22	110	4.6	2328	1	US-08-688-649-38
23	110	4.6	3499	1	US-07-966-278-2
24	110	4.6	3499	1	US-08-424-921-2
25	110	4.6	3499	2	US-08-556-355A-2
26	110	4.6	3499	2	US-07-803-627A-2
27	110	4.6	3499	4	US-09-244-889A-2

28	108.4	4.5	2328	4	US-09-715-524B-1	Sequence 1, Appli
29	106	4.4	5837	1	US-07-686-340-1	Sequence 1, Appli
30	106	4.4	5837	1	US-08-004-139B-1	Sequence 1, Appli
31	106	4.4	5837	1	US-08-117-491-1	Sequence 1, Appli
32	106	4.4	5837	1	US-08-271-364A-1	Sequence 1, Appli
33	106	4.4	5837	2	US-08-811-492-1	Sequence 1, Appli
34	106	4.4	5837	2	US-08-222-715B-1	Sequence 1, Appli
35	106	4.4	5837	5	PCT-US96-10545A-1	Sequence 1, Appli
36	102.6	4.3	1664976	4	US-08-916-421B-1	Sequence 1, Appli
37	99	4.1	4707	1	US-08-004-139B-2	Sequence 2, Appli
38	99	4.1	4707	2	US-08-811-492-2	Sequence 2, Appli
39	99	4.1	4707	5	PCT-US96-10545A-2	Sequence 2, Appli
40	87.4	3.6	2289	2	US-08-907-166-9	Sequence 9, Appli
41	87.4	3.6	2289	4	US-09-391-340-9	Sequence 9, Appli
42	77.6	3.2	2634	2	US-08-907-166-7	Sequence 7, Appli
43	77.6	3.2	2634	4	US-09-391-340-7	Sequence 7, Appli
c	44	62.2	7218	1	US-08-232-463-14	Sequence 14, Appli
45	56.2	2.3	1193	1	US-08-229-284A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-907-166-3
; Sequence 3, Application US/08907166
; Patent No. 5948666
; GENERAL INFORMATION:
; APPLICANT: Callen, Walter
; APPLICANT: Mather, Eric
; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
; FILE REFERENCE: 09010/027001
; CURRENT APPLICATION NUMBER: US/08/907,166
; CURRENT FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 2412
; TYPE: DNA
; ORGANISM: Pyrolobus fumarius
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2410)
US-08-907-166-3

Query Match		100.0%	Score 2412;	DB 2;	Length 2412;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 2412;		Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
Qy	1	ATGACTGAAGTGTGTTATTCACGGTTT	TAGACTCTAGCTACGAGGTGTTGGTAAAGAGCCT	60	
Db	1	ATGACTGAAGTGTGTTATTCACGGTTT	TAGACTCTAGCTACGAGGTGTTGGTAAAGAGCCT	60	
Qy	61	CAGGTAATCATATGGGTTATTCGTAGAA	CGGCGAGAGGTAGTCTCTATGACAGGTCT	120	
Db	61	CAGGTAATCATATGGGTTATTCGTAGAA	CGGCGAGAGGTAGTCTCTATGACAGGTCT	120	
Qy	121	TTTCGCCCATCTCTATGCGTGTGCTTG	CACCGGCGCGCTCCTTAAGCAGGTAGCACAA	180	
Db	121	TTTCGCCCATCTCTATGCGTGTGCTTG	CACCGGCGCGCTCCTTAAGCAGGTAGCACAA	180	
Qy	181	CGTATTCTGTCATTTAGTAGAGCCAA	GAGCCCGATTATAGGTGTAGAGATGACAAGAGG	240	
Db	181	CGTATTCTGTCATTTAGTAGAGCCAA	GAGCCCGATTATAGGTGTAGAGATGACAAGAGG	240	
Qy	241	AAGTACTTCGGAGGCGCTCGTAGGGTCT	TACGTATTCGACCGGTCTACCCGAGGCTGTT	300	
Db	241	AAGTACTTCGGAGGCGCTCGTAGGGTCT	TACGTATTCGACCGGTCTACCCGAGGCTGTT	300	
Qy	301	AGGAGTATCGCGAACTCGTAAAGACGT	TGATGGTGTGAGGATGTTCTAGAGGCGGAT	360	
Db	301	AGGAGTATCGCGAACTCGTAAAGACGT	TGATGGTGTGAGGATGTTCTAGAGGCGGAT	360	

QY	361	ATACGCTCCGCTATCTCATAGATCAAGATCTATTCTTTTCCCTTCCCTGATCCGT	420	1441	AGGTGTCCGCTCCGCTTCTTCAAGACAGTCTTCTTGAGAGCTCTTTAGAGCTTCTGTAAGCGT	1500
DB	361	ATACGCTCCGCTATCTCATAGATCAAGATCTATTCTTTTCCCTTCCCTGATCCGT	420	1501	GTGCGTGTCTGAATCAAGAGATCTCTCGGATACCCAGATATCTGATCTGATGAA	1560
QY	421	GTAGAGGTGAGCCCTCGAGAACAGATGGCTTCGGTGTGCAAGAGTATACCTGGTT	480	1501	GTGCGTGTCTGAATCAAGAGATCTCTCGGATACCCAGATATCTGATCTGATGAA	1560
DB	421	GTAGAGGTGAGCCCTCGAGAACAGATGGCTTCGGTGTGCAAGAGTATACCTGGTT	480	1561	AGCGAAGAGCGCTTGAAGGTTCTTCAAAACGCTAGTCTAGGCTACATGGTGTGAGCGGC	1620
QY	481	AAGAGAGCGCGAGCGACCTTTATGTGTAGGCTCTCGACCAACCAAGCTTCCGATCTT	540	1561	AGCGAAGAGCGCTTGAAGGTTCTTCAAAACGCTAGTCTAGGCTACATGGTGTGAGCGGC	1620
DB	481	AAGAGAGCGCGAGCGACCTTTATGTGTAGGCTCTCGACCAACCAAGCTTCCGATCTT	540	1621	GCTAGGTGTATGTGAGGAGTGTGCAAAAGGCTGTCAAGGCTTGGGTGAGGACCTCAT	1680
QY	541	AGGATACCTCGGTTTCAATTTATAGCAAGCAAGGCTCGCGCTTCCAGAGCGC	600	1621	GCTAGGTGTATGTGAGGAGTGTGCAAAAGGCTGTCAAGGCTTGGGTGAGGACCTCAT	1680
DB	541	AGGATACCTCGGTTTCAATTTATAGCAAGCAAGGCTCGCGCTTCCAGAGCGC	600	1681	CGCACCGCATCAACATAGCTCTGTAACCTAGGCTCAAGGTGATCTACGCTGACACAGAT	1740
QY	601	GATCTGTATAGTGTAGTGTGTAAGACTGACATGGCGATGAGGTGCTATTCAATTGCA	660	1681	CGCACCGCATCAACATAGCTCTGTAACCTAGGCTCAAGGTGATCTACGCTGACACAGAT	1740
DB	601	GATCTGTATAGTGTAGTGTGTAAGACTGACATGGCGATGAGGTGCTATTCAATTGCA	660	1741	TCGCTCTTCTGCTGACCTTATGATCCGAGAGAGGTGGAAATTTTCAATAAATTATAAGGAG	1800
QY	661	GAGGCAAGACGATCGAAACCGATACGCGAGTTGTAGAGTACGTGAGAGTATGAC	720	1741	TCGCTCTTCTGCTGACCTTATGATCCGAGAGAGGTGGAAATTTTCAATAAATTATAAGGAG	1800
DB	661	GAGGCAAGACGATCGAAACCGATACGCGAGTTGTAGAGTACGTGAGAGTATGAC	720	1801	GAGCTGGGTTTCAAAATCAAGCTAGAGAGGTGTACAAACGCTTATTTTACAGAGGCT	1860
QY	721	CCGACATAATAGTCCGTTTATAACCAATCAATTCGATTGGCTTATCTTTTTCAGGCGC	780	1801	GAGCTGGGTTTCAAAATCAAGCTAGAGAGGTGTACAAACGCTTATTTTACAGAGGCT	1860
DB	721	CCGACATAATAGTCCGTTTATAACCAATCAATTCGATTGGCTTATCTTTTTCAGGCGC	780	1861	AAGAGAGGTGTACGCTTCTCGAGGACGAGCTATAGATATTTCTCGGTTTCGAGGCT	1920
QY	781	GCCGCTATCTAGGATCAAGCTTGTATGACTAGAGAGTGTGGCGCGAGCCCAACT	840	1861	AAGAGAGGTGTACGCTTCTCGAGGACGAGCTATAGATATTTCTCGGTTTCGAGGCT	1920
DB	781	GCCGCTATCTAGGATCAAGCTTGTATGACTAGAGAGTGTGGCGCGAGCCCAACT	840	1921	GTACGTGGCGATTGTGTAATCTGCAAGAGGTTTCAAGCTTGTGAAATATGTA	1980
QY	841	ACGCTATCTAGGATCAAGCTTGTATGACTAGAGAGTGTGGCGCGAGCCCAACT	900	1921	GTACGTGGCGATTGTGTAATCTGCAAGAGGTTTCAAGCTTGTGAAATATGTA	1980
DB	841	ACGCTATCTAGGATCAAGCTTGTATGACTAGAGAGTGTGGCGCGAGCCCAACT	900	1981	TTGAGACGAGTGTGAGGTGTGTAATCTGCAAGAGGTTTCAAGCTTGTGAAATATGTA	2040
QY	901	GAGGATGTCAGAGATCAAGATAAGAGTCTCGAGAGGTTGCGAGATATCTAGGCGTG	960	1981	TTGAGACGAGTGTGAGGTGTGTAATCTGCAAGAGGTTTCAAGCTTGTGAAATATGTA	2040
DB	901	GAGGATGTCAGAGATCAAGATAAGAGTCTCGAGAGGTTGCGAGATATCTAGGCGTG	960	2041	GAGGAGGCAAGGTTCCATAGAGAGCTTGTAAATCTGGAAGACCTTTAGTAAGCGCTTT	2100
QY	961	ATGAGAGAGAGTGAACGCGTTTATCATCAATTTGGTGGAGATTTCAGAGATTTGGAGCAG	1020	2041	GAGGAGGCAAGGTTCCATAGAGAGCTTGTAAATCTGGAAGACCTTTAGTAAGCGCTTT	2100
DB	961	ATGAGAGAGAGTGAACGCGTTTATCATCAATTTGGTGGAGATTTCAGAGATTTGGAGCAG	1020	2101	GAGGAGTACAAACGAGGACCAACGCTGTCGAGGAGAGAGGATGCTGTGAGCAGGC	2160
QY	1021	CCGAGAGAGAGCCTATTAATGCAATACGCGCGAGATGTCGCGCTTACTTACGCGC	1080	2101	GAGGAGTACAAACGAGGACCAACGCTGTCGAGGAGAGGATGCTGTGAGCAGGC	2160
DB	1021	CCGAGAGAGAGCCTATTAATGCAATACGCGCGAGATGTCGCGCTTACTTACGCGC	1080	2161	TACCGGTAAACCCAGCGGACAGATAGGCTGTAAATAGTGAAGGTTGGTGGCGGTATC	2220
QY	1081	TTAGCCGAGAGATTTGCGCTTGTCTATCCAGTTGTCTGATCAACAGGTTCTCCACTA	1140	2161	TACCGGTAAACCCAGCGGACAGATAGGCTGTAAATAGTGAAGGTTGGTGGCGGTATC	2220
DB	1081	TTAGCCGAGAGATTTGCGCTTGTCTATCCAGTTGTCTGATCAACAGGTTCTCCACTA	1140	2221	AGTCAAGAGCATGGCCATCTTCAAGGATCTTAGCCAGATAGACGTGACCTAC	2280
QY	1141	GACCGAGTGTGAGTGTGCTTTCGATTTGATGATGATGATGATGATGATGATGATGAT	1200	2221	AGTCAAGAGCATGGCCATCTTCAAGGATCTTAGCCAGATAGACGTGACCTAC	2280
DB	1141	GACCGAGTGTGAGTGTGCTTTCGATTTGATGATGATGATGATGATGATGATGATGAT	1200	2281	TATTTGACCAACCAATCAATCCGCTGTGATGAGATACTGGGCTACTTTGGATCACC	2340
QY	1201	TTTAGATCAAGAGCTTGTGCGGACCGCTTGTAGCGCCAGAGAGATTTACCGTGGC	1260	2281	TATTTGACCAACCAATCAATCCGCTGTGATGAGATACTGGGCTACTTTGGATCACC	2340
DB	1201	TTTAGATCAAGAGCTTGTGCGGACCGCTTGTAGCGCCAGAGAGATTTACCGTGGC	1260	2341	GAGAGAGCTGTGAAGCAAGTGTGAACCTGGGCAAGAGCTCTCTTCGACTTTCTAGCCCAAG	2400
QY	1261	GCTATAGTCTTTGAGCGGTTGAGCGGCTGACAGAGATATAGCGGTACTCGACTTTAGC	1320	2341	GAGAGAGCTGTGAAGCAAGTGTGAACCTGGGCAAGAGCTCTCTTCGACTTTCTAGCCCAAG	2400
DB	1261	GCTATAGTCTTTGAGCGGTTGAGCGGCTGACAGAGATATAGCGGTACTCGACTTTAGC	1320	2401	AAGAGCAAGTAA 2412	
QY	1321	TCGATGTACCCCAACATCATGATAAAGTACATGTTGGTCTTGAACGCTTGTAGGCGCT	1380	2401	AAGAGCAAGTAA 2412	
DB	1321	TCGATGTACCCCAACATCATGATAAAGTACATGTTGGTCTTGAACGCTTGTAGGCGCT	1380			
QY	1381	GCTGAAAGTGTGGCGGTTGTGTTCTGAGAGGCTTTTGGAGGCTTTTGGAGGCTTTTGGAGGCT	1440			
DB	1381	GCTGAAAGTGTGGCGGTTGTGTTCTGAGAGGCTTTTGGAGGCTTTTGGAGGCTTTTGGAGGCT	1440			
QY	1441	AGGTGTCCGCTTCTTCAAGACAGTCTTTTGGAGGCTTTTGGAGGCTTTTGGAGGCTTTTGGAGGCT	1500			

RESULT 2
US-09-391-340-3
; Sequence 3, Application US/09391340A
; Patent No. 6492511
; GENERAL INFORMATION:

```
; APPLICANT: Callen, Walter
; APPLICANT: Mather, Eric
; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASSES
; FILE REFERENCE: 09010/027001
; CURRENT APPLICATION NUMBER: US/09/391.340A
; CURRENT FILING DATE: 1999-09-07
; EARLIER APPLICATION NUMBER: US 08/907,166
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 2412
; TYPE: DNA
; ORGANISM: Pyrolobus fumarius
; NAME/KEY: CDS
; LOCATION: (1) ... (2410)
US-09-391-340-3

Query Match      100.0%; Score 2412; DB 4; Length 2412;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGACTGAAGTGTATTCACGGTTTACACTCTAGCTACGAGGTGTTGGTAAAGACCT 60
Db |||
Qy 61 CAGGTAATCATATGGGTTATTCAGGTAACCGCGAGAGGAGTCTCATTCAGAGTCT 120
Db |||
Qy 121 TTTGCCCATCTTCTATGCGCTGCTGCACCGCGCGCCGCTCTTAGACGAGTACGACAA 180
Db |||
Qy 181 CGTATTCGTGCAATGAGTAGGCCAAAGAGCCGATATATAGGTGTAGAGGATGACAGAGG 240
Db |||
Qy 241 AAGTACTTCGGAGGCCCTCGTAGGTTTACGTTTACGTTTACGTTTACGTTTACGTTT 300
Db |||
Qy 301 AGGAGTATCCGAACTCGTAAGAACGTTGATGGTGTAGGATGTTCTAGAGCCGAT 360
Db |||
Qy 361 ATACGCTTCGCTATGCGCTATCTCATAGATCACGATCTATTTCTTTCACCTGTACCGT 420
Db |||
Qy 421 GTAGAGGCTGAGCCCTCGAGAACAGATGCGCTTCGTTGCGTACAGGATACCTGTT 480
Db |||
Qy 481 AAGAGCGCGCGAGCCACTTATGTTGAGGCTCTCGACCAACCAAGCTTCCGATCTT 540
Db |||
Qy 541 AGGATATCTCGGTTTCGATATTAAGTTTATAGCAAGAGGTCGCGCGTCCAGAGCGC 600
Db |||
Qy 601 GATCTCTGTAATGATGATGAGTGTGAAGCTGACGATGCGGATGAGGTGCTATTCATTGCA 660
Db |||
Qy 661 GAGGCGAAGACGATCGAAACCCGATACCGAGTTTGTAGATGACGTGAAGAGTATGAC 720
Db |||
Qy 721 CCCGACATAATAGTCGGTTATTAACCAATCAATTCGATTCGCTTATCTTTTACGAGCGC 780
Db |||
```

```
Db 721 CCCGACATAATAGTCGGTTATTAACCAATCAATTCGATTCGCTTATCTTTTACGAGCGC 780
Qy 781 GCCCGCATCTAGGCATAAAGCTTGATGTGACTAGAGAGTTGGCGCCGAGCCCACT 840
Db |||
Db 781 GCCCGCATCTAGGCATAAAGCTTGATGTGACTAGAGAGTTGGCGCCGAGCCCACT 840
Qy 841 AGGTCATATGGGCACGCTCTCTGCTCCCTGCGCGCTTAAAGTAGTCTGTACGATATGCC 900
Db |||
Db 841 AGGTCATATGGGCACGCTCTCTGCTCCCTGCGCGCTTAAAGTAGTCTGTACGATATGCC 900
Qy 901 GAAGAGATCCAGAGATCAAGATAAAGAGTCTCGAGAGGTCCGAGATATCTAGGCGTG 960
Db |||
Db 901 GAAGAGATCCAGAGATCAAGATAAAGAGTCTCGAGAGGTCCGAGATATCTAGGCGTG 960
Qy 961 ATGAAGAAGAGTGAACGCGTTATCATCAATTTGGTGGAGATTCAGAGCTATTTGGGACGAC 1020
Db |||
Db 961 ATGAAGAAGAGTGAACGCGTTATCATCAATTTGGTGGAGATTCAGAGCTATTTGGGACGAC 1020
Qy 1021 CCGAAGAAGAGACCACTATTACTGCAATACGCGCGACGATGTCGCGCTACTTACGCG 1080
Db |||
Db 1021 CCGAAGAAGAGACCACTATTACTGCAATACGCGCGACGATGTCGCGCTACTTACGCG 1080
Qy 1081 TTAGCCGAGAGATATTGCCGTTTGTATCCAGTTGTCTAGCTAACAGTCTCCCACTA 1140
Db |||
Db 1081 TTAGCCGAGAGATATTGCCGTTTGTATCCAGTTGTCTAGCTAACAGTCTCCCACTA 1140
Qy 1141 GACCAAGTATGAGAGAGTCTGTCGCGACCGCTTACGCGCCGAGAGAGTCTACCGTGCG 1200
Db |||
Db 1141 GACCAAGTATGAGAGAGTCTGTCGCGACCGCTTACGCGCCGAGAGAGTCTACCGTGCG 1200
Qy 1201 TTTAAGATGAAGAGAGTCTGTCGCGACCGCTTACGCGCCGAGAGAGTCTACCGTGCG 1260
Db |||
Db 1201 TTTAAGATGAAGAGAGTCTGTCGCGACCGCTTACGCGCCGAGAGAGTCTACCGTGCG 1260
Qy 1261 GCTATAGTTCTTTCAGCCGTTGAGAGCGGTGACAGAAATATAGCCGTCTCGACTTTAGC 1320
Db |||
Db 1261 GCTATAGTTCTTTCAGCCGTTGAGAGCGGTGACAGAAATATAGCCGTCTCGACTTTAGC 1320
Qy 1321 TCGATGTACCCAAACATCATGATAAAGTACAATTTGTTGCTTCTGACAGCTTTGTAGGCGCT 1380
Db |||
Db 1321 TCGATGTACCCAAACATCATGATAAAGTACAATTTGTTGCTTCTGACAGCTTTGTAGGCGCT 1380
Qy 1381 GGTGAAAAGTGTGCGAGTGTGTTGCTGCGAGGCGCCGAGGTCAAGACAGTTCCTT 1440
Db |||
Db 1381 GGTGAAAAGTGTGCGAGTGTGTTGCTGCGAGGCGCCGAGGTCAAGACAGTTCCTT 1440
Qy 1441 AGGTGTCCGCGCGCTTCTTCAAGACAGTCTTTCAGAGCGCTTTCAGAGCTTTCGTAAGCGT 1500
Db |||
Db 1441 AGGTGTCCGCGCGCTTCTTCAAGACAGTCTTTCAGAGCGCTTTCAGAGCTTTCGTAAGCGT 1500
Qy 1501 GTGCGTGTGAATAAGAAAGTATCTCCGATAGCCCAAGATATCGACTGTGGATGAA 1560
Db |||
Db 1501 GTGCGTGTGAATAAGAAAGTATCTCCGATAGCCCAAGATATCGACTGTGGATGAA 1560
Qy 1561 AGSCACAAGCGTTGAGGTTCTGCAACGCTAGTTCGCTAGCTAGCTAGCTAGCTAGCTAGCT 1620
Db |||
Db 1561 AGSCACAAGCGTTGAGGTTCTGCAACGCTAGTTCGCTAGCTAGCTAGCTAGCTAGCTAGCT 1620
Qy 1621 GCTAGTGTGTTTTCAGAGGAGTTCGCAAGGCTGTCACGCGTTGGGTTAGGCACTCAT 1680
Db |||
Db 1621 GCTAGTGTGTTTTCAGAGGAGTTCGCAAGGCTGTCACGCGTTGGGTTAGGCACTCAT 1680
Qy 1681 CGCACCGCATCAACATAGCTGTAACTAGGCTCAAGGTGATCTACGCTGACACAGAT 1740
Db |||
Db 1681 CGCACCGCATCAACATAGCTGTAACTAGGCTCAAGGTGATCTACGCTGACACAGAT 1740
Qy 1741 TCGCTCTTCTGACCTTATGATCCGAGAGAGTGGAAATTTTCATCAAAATTTAAGGAG 1800
Db |||
Db 1741 TCGCTCTTCTGACCTTATGATCCGAGAGAGTGGAAATTTTCATCAAAATTTAAGGAG 1800
Qy 1801 GAGCTGGGTTTCGAAATCAAGCTAGAGAGGTGTACAAAGCTTATCTTTTACAGAGCT 1860
Db |||
Db 1801 GAGCTGGGTTTCGAAATCAAGCTAGAGAGGTGTACAAAGCTTATCTTTTACAGAGCT 1860
```

1861 AAGAAGAGGTACGCTGGCTTCTTCGAGGACGACGATATAGATATTGTTCGGTTTCGAGGCT 1920
1861 AAGAAGAGGTACGCTGGCTTCTTCGAGGACGACGATATAGATATTGTTCGGTTTCGAGGCT 1920
1921 GTACGTGGCGATTTGGTGAACCTCGCAAGAGGTTTCAGACTAAAGTTTGTGCAATAGTA 1980
1921 GTACGTGGCGATTTGGTGAACCTCGCAAGAGGTTTCAGACTAAAGTTTGTGCAATAGTA 1980
1981 TTGACAGCAGGTGAGTGAACAGGCTGTAGAGTACCTCAGGAAGATTGTGGAAGTTG 2040
1981 TTGACAGCAGGTGAGTGAACAGGCTGTAGAGTACCTCAGGAAGATTGTGGAAGTTG 2040
2041 GAGGAGGCAAGGTTCCCATAGAGAGCTTGTAACTCGAAGACCTTGTAGTAAAGCTCTT 2100
2041 GAGGAGGCAAGGTTCCCATAGAGAGCTTGTAACTCGAAGACCTTGTAGTAAAGCTCTT 2100
2101 GAGGAGTACACAAAGAGGACACACGCTGTTGACGAGAGGATGCTGTACAGAGC 2160
2101 GAGGAGTACACAAAGAGGACACACGCTGTTGACGAGAGGATGCTGTACAGAGC 2160
2161 TACCGGTAAGCCAGGACGACAGATAGGATGATGTAATAGTGAAGGTTGGTGGCGTATC 2220
2161 TACCGGTAAGCCAGGACGACAGATAGGATGATGTAATAGTGAAGGTTGGTGGCGTATC 2220
2221 AGTCAAGAGCATGGCCATCTTTCATGTCAGAGATCCCTAGCCAGATAGACCTGACCTAC 2280
2221 AGTCAAGAGCATGGCCATCTTTCATGTCAGAGATCCCTAGCCAGATAGACCTGACCTAC 2280
2281 TATGTTGACACCAATCATCCGCTGATGAGATGATGAGATGATGAGATGATGAGATGAT 2340
2281 TATGTTGACACCAATCATCCGCTGATGAGATGATGAGATGATGAGATGATGAGATGAT 2340
2341 GAGAAGAGCTGAACCAAGTGCACCTGGGACAGAGCTCTCTCGACTTCTTACGCAAG 2400
2341 GAGAAGAGCTGAACCAAGTGCACCTGGGACAGAGCTCTCTCGACTTCTTACGCAAG 2400
2401 AAGAGCAAGTAA 2412
2401 AAGAGCAAGTAA 2412

RESULT 3
US-08-062-368-3
Sequence 3, Application US/08062368
Patent No. 5491086
GENERAL INFORMATION:
APPLICANT: Gelfand, David H.
APPLICANT: Wang, Alice M.
TITLE OF INVENTION: Purified Thermostable Nucleic Acid
TITLE OF INVENTION: Polymerases Enzyme From Pyrodicticum Species
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/062,368
FILING DATE: 19930514
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sias Ph.D., Stacey R.
REGISTRATION NUMBER: 32,630
REFERENCE/DOCKET NUMBER: 8584
TELECOMMUNICATION INFORMATION:

TELEPHONE: (510) 814-2863
TELEFAX: (510) 814-2977
INFORMATION FOR SEQ ID NO. 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2430 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-062-368-3
Query Match 44.3%; Score 1069.6; DB 1; Length 2430;
Best Local Similarity 66.5%; Pred. No. 0;
Matches 1601; Conservative 0; Mismatches 789; Indels 18; Gaps 4;
QY 11 TTGTATTACAGGTTTGTAGACTCTAGCTACGAGGTGTGTGTAAGAGCCCTCAGGTATCA 70
Db 14 TAGAGTTTCGTCTCTAGACTCTAGCTACGAGTACTCGGGAAGAGCCGCTAGTATCC 73
QY 71 TATGGGTATTGTGAGACGCGGAGAGGTTAGTCTCATTTGACAGGTCTTTTCGCCAT 130
Db 74 TCTGGGGATTAACCTTTGACGTTAACGTTGTCGTCTTTAGACCCCGCTTCGCCCT 133
QY 131 ACTTCTATGCGCTCTTTCACCGGCGCCGATCCCTA-----AGCAGGTAGCACACGTA 184
Db 134 ACTTCTACGCCCTCATAGCCCGGCTATGAGGATATGTTGAGGAGATAGCAGCTTCCA 193
QY 185 TTCTGTGATTGAGTAGGCAAGAGCCGATATTAGTGTAGAGATGACAGAGGAGT 244
Db 194 TAAGAGGCTTAGTGTGCTCAAGAGTCCGATAATAGATGCCAAGCCTTTGATAAGAGT 253
QY 245 ACTTCGGAGGCTTCGTAGGCTCTTACGTTTTCGACCGCTACCGGAGGCTTTAGGG 304
Db 254 ACTTCGGAGGCTTCGTAGGCTTCGAGGTTTACCATATGATACCGGATCTGTTAGAC 313
QY 305 AGTATCCGAACTCGTAAAGAACGTTGATGTTGAGGATGTTCTAGAGCGGATATAC 364
Db 314 ACTACCGGAGGCTGAAAGAGATAGAGGTTGAGGAGTCTCCCTCGAGGAGATATAA 373
QY 365 GCTTCGCTATGCGCTATCTATAGATCAGATCATTTCTTCTTCCCTACCTGTTACGTTAG 424
Db 374 GGTTCGAATAGATATCTGATAGATAGAGGCTCTACCGGTTTACCGGATCC 433
QY 425 AGGCTGAGCCCTCGAGAACAAAGATGGCTTTCGTTGTCGACAAAGGTATCTCTGTTAAGA 484
Db 434 CCGTAGAGGATGCGGGCGGCTATCCAGGCTTCGTTGTTGACCGTGTCTACAAGGTTCTG 493
QY 485 GCAGCCGAGGCTCACTTTATGTTGAGGCTCTGCGACCAACCAAGCTTCCGATCTTAGGA 544
Db 494 GCGACCGGAGCCCT-----AGCGGATATAACCGGATCGACCTTCCCGGATGAGGC 547
QY 545 TACTCGCTTCGATATTGAAAGTTTATAGCAAGCAAGGTTGCGCGCTCCGAGCGGATC 604
Db 548 TGGTAGCTTTTGTATATAGAGTGTATAGAGGAGGTTTCTGAGGAGTCTGAGAGAGGATC 607
QY 605 CTGTAATAGTATAGCTGTGAAGACTGACGATGCGGATGAGGTTGCTTATTCATTCAGAGG 664
Db 608 CAGTGATAATAGTGTCTGAGGACAGCGGAGGAGAGGCTCATAGAGGCTGAAG 667
QY 665 GCAAGAGCATCGAAACCCGATACGCGAGTTTGTAGAGTACGTGAAGAGGATGAGCCCG 724
Db 668 GCCATGACGACAGAGGCTTCTGAGGAGTTCGTAGAGTACGTGAGAGAGGCTTCCGACCC 727
QY 725 ACATAATAGTGGTTTAAACAACTCATTTTCGATTTGGCTTATCTTTTTCGAGCGGCGCC 784
Db 728 ACATAATAGTGGCTTAAACAGTAAACCTTCGACTGGCCCTTACCTTAATGAGCGGCGCC 787
QY 785 GCATCTTAGGCATAAAGCTTTGATGCTAGTGAAGAGTTGGCGCGGAGCCACCATAGCG 844
Db 788 GTAGGCTCGGATTAACCTCGAGTTACACCGCTGTGCGGAGAGGCTCCACCAAGCG 847
QY 845 TACATGGGACGCTCTGTCTCCCTGGCAGGCTTAAAGGTAGATCTGTAGACTATGCGGAAG 904

848 TCTACGCCCACTCTCGGTGACGGTAGGCTGAACGTGACCTCTACGACTATGCGAGG 907
QY
905 AGATGCCAGAGTCAAGATTAAGAGTCTGAGAGGTCCAGAGTATCTAGCGGTATGA 964
Db
908 AGATGCCAGAGTAAAGATGAAGACCTTGAAGAGGTAGCGAGTACCTTAGCGGTATGA 967
QY
965 AGAAGAGTGAACCGGTATCATCAATGGTGGAGATTCCAGACTATTGGGACGACCGA 1024
Db
968 AGAAGAGCGAGCGTGTATATAGAGTGTGGAGNATACCGAGTACTGGGATGACGAG 1027
QY
1025 AGAAGAGACATATTACTGCAATACCGCGGAGAGATGTCGCGCTACTTACGCTTAG 1084
Db
1028 AGAAGAGGAGCTGTAGAGCGCTACGCGCTGACGATGTAGGCGTACTTACGCGCTCG 1087
QY
1085 CCGAGAGATATTGCGGTTTGTCTATCCAGTGTCTGATACAGAGTCTCCACATAGACC 1144
Db
1088 CGAAGAAGATGCTACCGTTTGGCCATACAGCTCTCCACTGTTACGGGTGTGCTCTCGACC 1147
QY
1145 AGTAGGTGCGATGAGTGTGGCTTTCGACTTGAATGGTACCTGATACCGCGGGGTTTA 1204
Db
1148 AGTAGGTGCTATGGCGTAGGCTTCCGCTAGAGTGTATCTCATGCGTGACGCTAGG 1207
QY
1205 AGATGAAGAGCTTGTGCGAAGCGGTTGAGCGGCCAGAGAGACTTACCGTGGCGTGA 1264
Db
1208 ATATGAACGAGCTGTGTGCGAAGCGGTTGAGAGGAGGGGGGAGAGCTACAAGGGTGCAG 1267
QY
1265 TAGTCTTCTGAGCGGTTGAGAGCGGTGACAGAGNATATAGCCGTACTCGACTTTAGCTGA 1324
Db
1268 TAGTGTAAAGCCTCTCAAGGAGTCCATGGAATGTTGTGGTCTCGAATTTCTAGTTCCA 1327
QY
1325 TGTACCCAAACATCATGATAAAGTACAATGTTGGTCTGACACGCTTGTGAGCGCTGCTG 1384
Db
1328 TGTACCCGAGCATATGATAAAGTACAACGTGGGCCCGACACTATAGTTCGACGACCCCT 1387
QY
1385 AAAAGTGTGCGAGT---GTGGTGTGGAGGCGCCCGGAGGTCAAGCACAGGTTCCGTA 1441
Db
1388 CGAGGTGCCAAAGTACGCGCGCTGTATGTAGTACCGCCGAGGTGCGGCGACCGGTTCCGTC 1447
QY
1442 GTGTGCGCGCGCTCTTCAAGACAGTCTTGTGAGAGGCTGTAGAGCTTGTGAGCGTG 1501
Db
1448 GCTCCCGCGCGCTCTTCAAGACAGTCTTGTGAGAGGCTGTAGAGCTTGTGAGCGTG 1507
QY
1502 TCGGTCTGAAATGAAGAAGTATCTTCCGATAGCCGAGATATCGACTTGTGGATGAAA 1561
Db
1508 TAAAGGAGAAGTGAAGGAGTTCCGCTGACAGCGCCGAGTACAGGCTTACGATGAGC 1567
QY
1562 GCGAGAGCGGTTGAAGGTTCTGMAACGCTAGTTTACGCTACATGGTTGAGAGCGGCG 1621
Db
1568 GCGAGAGCGGCTCAAGGTTCTTGGAGACGCGAGCTATGGCTACATGGGTTGAGGCGCATG 1627
QY
1622 CTAGGTGGTATTGCGAGGAGTGGCAAGGCTGTACGCTTGGGTTAGGCACTCATAC 1681
Db
1628 CCGGTGTGTTGCAACCGCTGCGCGAGGCTGTACAGCTTGGGCGGTAACTTATAC 1687
QY
1682 GACCGCCATCAACATAGCTCGTAACTAGGCTCAAGGTGATCTACGCTGACACAGATT 1741
Db
1688 TGACAGCTATCGAGTATGCGAGAGCTCGGCTTAAGCTTATATGAGAGACACCGACT 1747
QY
1742 CCGCTCTGTTGACCTATGATCGGAGAGGTTGGAATAATTCATCAAAATATAAAGGAGG 1801
Db
1748 CCGCTCTGTTGTTATGATCAAGGAGAGGTTGAGAGCTGATAGAGTTTGTGAGAGG 1807
QY
1802 AGCTGGGTTTCAAACTCAAGCTAGAGAGGTTGATCAAAAGCTTATCTTACAGAGGCTA 1861
Db
1808 AGCTGGGCTTGGATTAAGATAGACAAAGTCTTCAAGAAAGTGTCTTACGGAGGCTA 1867
QY
1862 AGAAGAGTACGCTGCGCTTCTCGAGGAGCGAGTATAGATATTGTCGGTTTCGAGGCTG 1921
Db
1868 AGAAGCGCTATGAGTCTCTCGAGGAGCGAGCTATAGACATCGTGGCTTTGAGGAG 1927
QY
1922 TAGTGGCGAATGTTGTTGAACTCGCCAGAGGTTTCAGACTAAGGTTGTGAAATAGTAT 1981
Db
1928 TCCGCGGCGACTGGTTCGAGCTGGCTAAGGAGTGCAGAGAGGCGGCTGAGATAGTGT 1987

1982 TGAAGCAGCTGAGGTGAACAGCGCTGTAGAGTACGTCAGGAAGATTGTGTAAGAGTTGG 2041
Db
1988 TGAATACGGGGAACGTGGACCAAGGCTATAAGCTACATAAGGAGGTAATAAAGCAGCTCC 2047
QY
2042 AGGAGGGCAAGGTTCCCATAGAGAAGCTTGTAAATCTGGAAGACCCCTTAGTAAGCGTCTTG 2101
Db
2048 GCGAGGGCAGGTCCTCAATACAAAGCTTATCATATGGAAGACCCCTGAGCAAGGATAG 2107
QY
2102 AGGATGACACAACCGAGGACACACACCTGCTGTCAGCGAAGAGAGATGCTGTACAGCGCT 2161
Db
2108 AGGATGACAGCATACGCGCTCATGTGTGCTGCTGACGCGCTATGAAGGAGGCGAGCT 2167
QY
2162 ACCGGTAAGCCAGGCGGACAGATAGGATGTATGATGAGAGGTTGGTGGCCGATACA 2221
Db
2168 ACGAGGTGTCTCCCGCGGATAGGTTGGCTACGTCATGATTAAGGAGTGGGAGTGTGT 2227
QY
2222 GTCAAGAGCATGCGCATCTTCAAGTGTCAAGGATCCTAGCCAGATAGAGTACCTACT 2281
Db
2228 CCAGCAGGCTTACCCCTACTTCAAGT---TGATCCATCGACCATCGATCTACTACT 2284
QY
2282 ATGTTGACCAACCAATCATCCCGCTGCATTGAGATGACTGGGCTACTTTGGCATCACCG 2341
Db
2285 ATATTGACCAACCATAGTGCAGCTGCTCTGAGGATCTCTCTACTTCGAGTCAACG 2344
QY
2342 AGAAGAGCTGAAGGCAAGTGCACACTGGCGAGAGACTCTTTCGACTTTCAGCCAGA 2401
Db
2345 AGAAGAGCTCAAGCGCGGCTACGTTGCGAGAGAGGCTCTTTCGACTTCTTGGCTCAA 2404
QY
2402 AGAGCAAG 2409
Db
2405 AGAATAG 2412

RESULT 4
US-08-062-368-1
; Sequence 1, Application US/08062368
; Patent No. 5491086
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Wang, Alice M.
; TITLE OF INVENTION: Purified Thermostable Nucleic Acid
; TITLE OF INVENTION: Polymerases Enzyme From Pyrodicticum Species
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/062,368
; APPLICATION NUMBER: 19930514
; FILING DATE: 19930514
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sias Ph.D. Stacey R.
; REGISTRATION NUMBER: 32,630
; REFERENCE/DOCKET NUMBER: 8584
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2863
; TELEFAX: (510) 814-2977
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2430 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

i: MOLECULE TYPE: DNA (genomic)

US-08-062-368-1

Query Match 43.3%; Score 1045.6; DB 1; Length 2430;
Best Local Similarity 65.9%; Pred. No. 0;
Matches 1566; Conservative 0; Mismatches 804; Indels 18; Gaps 4;

QY 11 TTGATATCAGCGTTTACCTAGCTACGAGGTTGTTGTAAGAGCCCTCAGGTATCA 70
DB 14 TAGAGTTTGTCTCTTGTATCAAGCTACGAGATTTAGGAAAGAGCCCGTATCATAC 73
QY 71 TATGGGTATTTGCTGAGAAACGCGAGAGGTTAGTCTCATGACAGTCTTTTCGCCCAT 130
DB 74 TATGGGTATTTGCTGAGAAACGCGAGAGGTTAGTCTCATGACAGTCTTTTCGCCCAT 133
QY 131 ACTTCTATGCGCTGCTTCCACCGCGCCGATCCCTA-----AGCAGTAGCACACGTA 184
DB 134 ACTTCTATGCGCTGCTTCCACCGCGCCGATCCCTA-----AGCAGTAGCACACGTA 193
QY 185 TTCGTGCAATTGAGTAGGCCCAAGAGCCCGATTTATAGGTGTAGAGGATGACAAGAGAAAT 244
DB 194 TTAGAAGGCTAAGTATGCAAGAGCCCATTAAGAGGCAAGGTTGTTAGTAAGAAGT 253
QY 245 ACTTCGGAGGCTCGTAGGCTCTTACGATTCGACCGTCTACCGAGGCTTTTAGGG 304
DB 254 ACTTCGGAGGCTCGTAGGCTCTTACGATTCGACCGTCTACCGAGGCTTTTAGGG 313
QY 305 AGTATCGGCACTGTAAGAAACGTTGATGTTGTTGAGGATGTTCTAGAGCGCGATATAC 364
DB 314 AATATAGAGGCTGTAAGAAACGTTGATGTTGTTGAGGATGTTCTAGAGCGCGATATAC 373
QY 365 GCTTCGCTATGCGCTATCTCATGATCAGGATCTATTTCTTCCACCTGCTAGCGGTAG 424
DB 374 GCTTCGCTATGCGCTATCTCATGATCAGGATCTATTTCTTCCACCTGCTAGCGGTAG 433
QY 425 AGGCTGAGCCCTCGAGAACAGATGGCTTCCGCTGTCGACAGGATATCTGTTAAGA 484
DB 434 GAGCGGAGAACGCTGGACGAGCCCTGTTTCCGTTGAGACTCGGTATAC-----ACTA 487
QY 485 CGAGCCGAGGACCTTTATGGTGGGCTCTCGACCAACAGCTTCCCGATCTTAGGA 544
DB 488 TAGTTGAGGACCGAGGCTATTGCGGACATAAAGTAGTATAGATATACAGAGATGCGTG 547
QY 545 TACTCGCTCGATATTGAGCTTTATAGCAGAGGAGGTCGCGGTCGAGAGGCGATC 604
DB 548 TCTTCGCTCGATATTGAGCTTTATAGCAGAGGAGGTCGCGGTCGAGAGGCGATC 607
QY 605 CTGTAATAGTATGCTGTGAAGACTGACGATGCGATGAGGTTCTATTCAATTCGAGAGG 664
DB 608 CGGTCAATAATCTCGATAAAGGACAGCAAGGCGGACGAGAGCTTACTAGAGGCAATA 667
QY 665 GCAGAGCGATCGAAGACCGATACCGAGTTTGTAGAGTACGTTGAGAGGATGACCCCG 724
DB 668 ACTAGCAGCAGAGAACGCTGCTACGGGAAATTTATAGAGTACATACGCTTCTTGAACCCAG 727
QY 725 ACATAATAGTCTGTTTATACCAATCATTTTGGATGCGCTTATCTTTTGGAGCGGCGCC 784
DB 728 ACATAATAGTCTGTTTATACCAATCATTTTGGATGCGCTTATCTTTTGGAGCGGCGCC 787
QY 785 GCATCCTAGGATCAAGATTTGATGACTAGAGAGTTGGCGCGAGCCCAACACTAGCG 844
DB 788 ACAGAAATAGGATTAAGCTGACGCTGACAGGCGGTTTGGCGGAGAGCCCAAGTATGAGCG 847
QY 845 TACATGGGCGAGCTCTGCTGCTGCGAGGCTTAACGTAGATCTGACGCTATGCGCGAG 904
DB 848 TCTATGACATGCTCTGCTGCGAGGCTTAACGTAGATCTGACGCTATGCGCGAG 907
QY 905 AGATCCGAGATCAAGATTTGATGACTAGAGAGTTGGCGCGAGCCCAACACTAGCG 964
DB 908 AATGCGATGAGATAAAGCTTAAAGACGCTCGAGGAGGTCGCGGAAATACCTAGGCGTTATGC 967
QY 965 AGAAGGCTGACGCGCTTATCATCTGTTGGTGGAGATTCAGACTTTGGAGCGGCCGA 1024

DB 968 CCAAGGCGAGCGGCTACTATAGATGTTGGTGGGATCCAGATTTACTGGACGCGAGA 1027
QY 1025 AGAAGAGACCACTATTACTCTCAATACGCGCGCGAGATGTCGCGCTACTTACGCGCTTAG 1084
DB 1028 AGAAGCGCGCTACTGAGCGGTTATGCGCTTCGACGATGTGAGAGCCACTTACGCGCTCG 1087
QY 1085 CCGAGAGATATTTCCCGTTTGTATCCAGTTGTGCTAGCTAAACAGAGTCTCCACACTAGACC 1144
DB 1088 CCGAGAGATATTTCCCGTTTGTATCCAGTTGTGCTAGCTAAACAGAGTCTCCACACTAGACC 1147
QY 1145 AGTAGGTGCGATGAGTGTGTTGCTTTCGACTTGAATGTTGTTGCTTACGCGCGGCTTTA 1204
DB 1148 AAGTCGCGGCTATGCGGCTAGGTTTCCGCTCTAGATGTTGTTGCTTATGAGAGAGCGCATG 1207
QY 1205 AGATGAAGAGCTTGTGCGCAACCGGTTGAGCGCCCGAGAGAGCTTACCGTGGCGCTA 1264
DB 1208 ATATGACGAGCTTGTCCCAACCGTGTCAAGCGCGCGAGAGAGCTTACAAAGGAGCAG 1267
QY 1265 TAGTTCTTGAGCCGTTTGAGAGGCGTGCACGAGAAATATAGCCGTTACTTCGACTTTAGCTCGA 1324
DB 1268 TAGTACTAAGCCCGCTTAAAGGCTTCCATGAGAAACGTAGTAGTGTCTGACTTTAGCTCAA 1327
QY 1325 TGTACCCAAACATCATGATTAAGTACGATGTTGTTGCTGTCGACGCTTGTGAGGCTGCGTG 1384
DB 1328 TGTACCCAAACATGATTAAGTACGATGTTGTTGCTGTCGACGCTTGTGAGGCTGCGTG 1387
QY 1385 AAAAGTGTGCGAGT---GTGTTGCTGGGAGCGCCCGAGGTCACAGCAGCTTCCGTA 1441
DB 1388 CAGAGTCGAGAGTACAGTGGATGCTACGTAGCCCGGAGTCGCGCACATGTTTAGGC 1447
QY 1442 GGTGTCGCGCGGCTTCTTCAAGACAGTCTTTCGAGAGGCTGTTTAGAGCTTCTGTAAGCGTG 1501
DB 1448 GCTCGGCTCGGCTTCTTAAAGACCGTGTCTGAGAACCTCTATAGCGCTGCGTAAAGCAAG 1507
QY 1502 TCGGTGCTGAAATGAAGAGTATCTCCGATAGCCAGAAATATCGACTGTTGCGATGAA 1561
DB 1508 TACGTGAAGAGTGAAGAGTCTCCCGCAGATAGCCGAGATACCGGATATACGATGAC 1567
QY 1562 GCGAGAGGCGTTGAAAGGTTCTTGAACACCTAGTTTACGCTACATGGGTTGAGCGCGCG 1621
DB 1568 GCCAAGAGGCTCAAGCTGTAGCCAAACGCTAGCTACGCTACATGGGATGCGTACAG 1627
QY 1622 CTAGTGTGATTTGAGGAGTGCAGAGGCTGCAAGGCTGTCAGCGCTTGGGTTGAGCGCTATAC 1681
DB 1628 CTGCTGTGATCTGTAACGCTGCGAGGCTGTAAAGCTTGGGCGCTTAAAGCTGATAC 1687
QY 1682 GCACCGCATCAACATAGCTCGTAAACTAGGCTTCAAGGTGATCTACGCTGACACAGATT 1741
DB 1688 TCTAGCAATAGATATGCTAGGAGCTCGGCTCAAGATATATATACGAGACACGACT 1747
QY 1742 CGCTCTTGTGACCTATGATCCGAGGAGGTTGAGAAATTTCAATCAAAATATATAAGGAGG 1801
DB 1748 CCTATTCTGTAACCTATGATATGAGAGGTTAAGAGCTTAAAGATTTAGATTTGCTGAGAAAC 1807
QY 1802 AGCTGGGCTTGAATCAAGCTAGAGAGGTTGACAAAGCTTATTTCTTTTACAGAGCTTA 1861
DB 1808 AGCTAGGCTTGAAGTAAAGATAGACAGGTTATACAAAGAGTGTCTTTTACCGAGCAA 1867
QY 1862 AGAAGAGTACGCTGCGCTTCTGAGAGCGGATGATAGATTTGCTGCTTTCGAGGCTG 1921
DB 1868 AGAAGAGTACGCTGCGCTTCTGAGAGCGGCTTTCGAGACATGATGAGGCTTTCGAGGCTG 1927
QY 1922 TACGTGGCGATTTGTTGAACTCGCAAGGAGGTTTCAGACTAAGTGTGCGAAATATAGTAT 1981
DB 1928 TTAGGCGGACTGTTGAGTATGAGAGGTTGAGAGGTTGAGAGGTTGAGAGTATATAC 1987
QY 1982 TGAAGAGAGTGGTGAACAGGCTGTAGAGTACGTCAGAGAGATTTGTAAGAGGTTGG 2041
DB 1988 TGAAGAGGAGAGATATAAGAGCTTACATAAGAGAGGTCGTTGAGAGAGCTAA 2047
QY 2042 AGGAGGCGAGGTTCCCATAGAGAGGTTTCTATCTGAGAGAGCTTGTAGTACGCTTTCG 2101
DB 2048 GAGAGGCGAGTATACCATTAACAAAGCTCGTAAATATATGGAAGACCTTTGACAAAGAGTACG 2107

Db 1826 ATGGAAGAAATATTTGTAAGGCTTCTGAAGTCAGAGAGCGGACTGGTGGAGCTTGCA 1885
 QY 1949 AGGAGGTTACAGCTTAAGGTTGTCGAAATAGTATTTGAAGAGGAGTGAAGTGAACAGGCTG 2008
 Db 1886 AGAAGATACAGAAAGGTTGTAAGAAATCATCTGAAGGAAAGAAATCTCTGAAAGAGCTG 1945
 QY 2009 TAGAGTACGTCAGGAAGATTTGTAAGAGTTGGAGGAGGCGCAAGGTTCCCATAGAGAGC 2068
 Db 1946 CTGAGTATGTAAGAGGAGTCATAGAGAGATTAAGGCGAGCAAAATTCGCTTGAGATT 2005
 QY 2069 TTGTAATCTGAAGACCTTAGTAAAGCTCTTGAGGAGTACAAACGAGGACACACAG 2128
 Db 2006 ATATCATCTACAGGAGTTGACGAGAAACCATCAAGTACGAGAGTATGACAGGCTCAG 2065
 QY 2129 TCGTTGACGCAAGAGAGATGTCGACAGGCTACCGGTAAAGCCAGGCGAGCAAGATAG 2188
 Db 2066 TAAAGCTGCGCATGAGCGCGCAAGAGAGATAGTATACAAATCGGCTCAAGGTTG 2125
 QY 2189 GGTATGTAATAGTGAAGGTTGGTGGCGGTATCATAGTCAAGAGAGATGGCCATCTTCATGG 2248
 Db 2126 GTTTTGTGTTTACAAAGGTTGTTGGGAACATAGTGTATAGGGCTTTTCCATCTGATCTGA 2185
 QY 2249 TCAAGGA 2255
 Db 2186 TAGAGGA 2192

RESULT 6

US-09-391-340-5
 ; Sequence 5, Application US/09391340A
 ; Patent No. 6492511
 ; GENERAL INFORMATION:
 ; APPLICANT: Callien, Walter
 ; APPLICANT: Mather, Eric
 ; FILE OF INVENTION: 09010/027001
 ; CURRENT APPLICATION NUMBER: US/09/391.340A
 ; CURRENT FILING DATE: 1999-09-07
 ; EARLIER APPLICATION NUMBER: US 08/907,166
 ; EARLIER FILING DATE: 1997-08-06
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 5
 ; LENGTH: 2367
 ; TYPE: DNA
 ; ORGANISM: Archaeoglobus lithotrophicus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(2364)
 US-09-391-340-5

Query Match 11.6%; Score 279; DB 4; Length 2367;

Best Local Similarity 50.9%; Pred. No. S.le-78; Mismatches 795; Indels 63; Gaps 7;

Matches 889; Conservative 0;

QY 530 TTCGGATCTTAGGATACCTCGGTTTCGATATTTGAAGTTTATAGCAAGCAAGGTCGCCGC 589
 Db 488 TTCGGAATCAAGTTATGCGTTTGAATGCGAATGCTCTCAGAGGTTGGTATGCCG 547
 QY 590 GTCCAGAGCGGATCTCTGTAATAGTATGATGTGAAGTACGATGAGGCGATGAGGTC 649
 Db 548 ATCCAGAGAAAGATCCCTATCATATGATCATATCAATTAATCGGTTGAATACGAGAAATCC 607
 QY 650 TATTCAATGACAGGCGCAAGACGATCGAAACCGATACCGGATTTGTAGATACGTGA 709
 Db 608 TCAAGGTTG-----ATACAGAGAGAGATTTGCTTACCAGATTTGTCAAGATATTC 658
 QY 710 AGAGTATGACCCGACATATAGTCGTTTATACAAATCATTTTCGATGTCCTTATC 769
 Db 659 GCGATATGATCCCGACATATATAGTTGATACATACAGGACGTTTTCAGTGGCCCTATA 718
 QY 770 TTTTGAAGCGCGCGGATCTCTAGCATTAAGCTTTGATGATAGAGGTTGGCGCG 829

Db 719 TCAAGAGAGAGCTGAGAACTAGAGGTTAAGCTTGACATCGGAAGAGTGAAGCGAAC 778
 QY 830 AGCCACCACTAGAGTACATGGGCACGCTCTGTCTCCCTGGCAGAGCTTAAACGTAGATCTGT 889
 Db 779 TGGCTATCAGGGAGGAGAACCCAAA-----GATTGCTGGCAGGTTGAACGTGGATCTCT 832
 QY 890 ACAGTATGCGGAAAGAGATGCCAGAGATCAAGATAAAGAGTCTCGAGGAGGTCGACAGAT 949
 Db 833 ATGATATGCAATGAGGAGTCTCGATGTAAGGTGAAGAGCTCGAAACGTTGACAGAT 892
 QY 950 ATCTAGCGGTGATGAAGAGAGTGAACGGGTATATCAATTTGGTGGGAGATTCAGACT 1009
 Db 893 TTCTGGTAAAGAAATAGAGCTTTCAGATATTTGAAGCGAAGGATATCTACAAGCACTGGA 952
 QY 1010 ATTGGGACGACCCGAAAGAGAGAGACCACTATTACTGCAATACCGCGCGGAGTATCCGCG 1069
 Db 953 CATCGGCGACAGGAAAGCG-----TAATCAAAATACCTCCCGCAGGACATCTCTGCG 1003
 QY 1070 CTACTTACGGCTTAGCGGAGAGATATTGCCGTTTGTCTATCCAGTTTGTCTGCTAAGACAG 1129
 Db 1004 ACAGTACTTTCATAGCTGAAGATTTGCTGCCAATGCTATTCGAACTTTCCAGATATAC 1063
 QY 1130 GTCTCCCACTAGACACGAGTGAAGTGTGCGATGAGTGTGCTTTCGACTGTAATGCTACCTGA 1189
 Db 1064 GCATACCTCTCGATGATGTGACAAAGAGCGGAGAGGTAAGCAGGTTGATGCTGCTGT 1123
 QY 1190 TAGCGCGCGGCTTTAAGATGAAAGAGCTTGTGCCGAACCGCTTGAGCCGCCAGAGAGA 1249
 Db 1124 TAAGCGAAGCACAAACTTTGGCGAACTTTGCACCCCAACCCAGAGAG---ATGSCCGACA 1180
 QY 1250 CTTACCGTGGCGCTATAGTTCTTGACCGTTGAGAGCGGTGACAGGAAATATAGCGGTAC 1309
 Db 1181 GCTATGAAGAGGATTCGTGCTCGGCCCGCAGAGAGGATTTGATGAGAACGTAATCTGCC 1240
 QY 1310 TCGACTTTAGCTCGATGTAACCAACATCATGATAAAGTACAAATTTGCTCTGCAACGC 1369
 Db 1241 TGGACTTTGCTCCATGATCTCCCTCAATATGATTTTCATACACATCAGCCCGCAGACGC 1300
 QY 1370 TTGTGAGCGCTGTTGAAGAGTGTGGCGAGTGTGTTGCTGGAGGCGCCGAGGTTCAAGC 1429
 Db 1301 TTGT-----AATAGGCAAAATGCGACGATTTGCAATCTAGCCCGGAGGTTGGGCG 1348
 QY 1430 ACAGGTTCCGTAGGTGTCGCCCGCTTCTTCAAGACAGTTCTTGAAGGCTGTTAGAGC 1489
 Db 1349 ACAAATTCAGGAACATCTCTGATGTTTTCATGAAGATATCTCAAAATGCTGATGAGA 1408
 QY 1490 TTGTAAGCGTGTGCTGTGAATGAAGAGTATCTCCGGATACCCGAGAAATATCGAC 1549
 Db 1409 AAAGAGAGAAATAAAGAGGTTATGAAACACACTTGACTACAACTCGCCAGATACNAGC 1468
 QY 1550 TGTGATGAAGGAGGAGGCGCTTGAAGGTTCTTCAAGGCTTCTGCAACCGTAGTTACGCTACATGG 1609
 Db 1469 TGCTGATATTAAGCAGGCGAACGCTGAAAGTTTCTTACAACTGTTTACGGTTTATATG 1528
 QY 1610 GTTGGAGCGCGCTAGGTGGTATTGAGGAGTGTGCGCAAGGCTGTGACCGGTTGGGGTA 1669
 Db 1529 GGTGGAGTCTTTCGAGATGTTACTGCAAGGAGTGTGCTGAAGCTACAAAGGATGCGGCA 1588
 QY 1670 GGCACCTCATACGCGCGCATCAATAGTCTGTAACATAGGCTCAAGGTGATCTAGC 1729
 Db 1589 GACACTTTATCAAAACATCTGCAAGAAATTCGCAAGAGGTTGGATTTGAAGTGTCTATG 1648
 QY 1730 GTGACACAGATTCGCTCTTCGTGACCTCATGATTCGGAGAGGTTGGAATAATTTCACTAAA 1789
 Db 1649 GGGATACAGATAGCATCTTTGTTAABAAGATGGATTTGAGCTGGAGAGCTCAAAAG 1708
 QY 1790 TTATAAA-----GGAGAGCTGGGGTTGGAATTCGAAGTCAAGCTAGAGA 1828
 Db 1709 AAGTTAAAGAGCTCATAGGTAAACTTTCCGAGAGAGATGCCAATACAAATAGAGATAGATG 1768
 QY 1829 AGGTGTACAAACGCTTATCTTTACAGAGGCTTAAGAGAGTACCTGCGCTTCTTCGAGG 1888
 Db 1769 AATCTACGAGACATATCTTCGTTGAA---AAGAAAGGTTATGCTGGATGACACAGG 1825

```

QY 1889 ACGGACGTATAGATATTCGCTTCGAGGCTGTACGTCGAGGATGGTGTGAATCCGCCA 1948
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1826 ATGGAAGATAATTCGTAAGGCTCTTGAAGTCAGAAAGAGCGACTGGTGCAGCTTGCAA 1885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1949 AGGAGTTCAGACTAAGCTTCGAAATAGTATTGAACAGAGTGTGAGTGAACAGCTG 2008
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1886 AGAAGATACAGAAAGGTGTAAATAGAAATCATCTCAAGGAAAGAAATCCTGAAAGAGCTG 1945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2009 TAGAGTACGTACAGGAGATTTGTAAAGAGTTTGGAGGAGCGCAGGTTCCCATAGAGAGC 2068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1946 CTGAGTATGTGAAGGAGTATAGAGGAGATAAAGCGCAGGCAAAATTCGGCTTGAAGATT 2005
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2069 TTGTAATCTGGAGACCCCTTAGTAAGCGCTTGTGAGGATACACACGAGGAGCCACACG 2128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2006 ATATCATCTACAAAGGATTTGACGAGAAACCATCAAGTACGAGATGATGAGGCTCACG 2065
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2129 TCGTTGACGAGAGGAGTGTCTGACGAGGCTACCGGTTAAGCCGACAGATAG 2188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2065 TAAAGCTGCCATGAGCGGCAAGAGAGGATATATACATCGGCTCAAGGTTG 2125
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2189 GGTATGTAATAGTGAAGGTTGTGCGGCTATCAGTCAAGAGAGTATGCGCATCTTCATGG 2248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2126 GTTTGTCTTACAAAGGTGTGGGAAACATAGGTATAGGCTTTTCCATCTGATCTGA 2185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2249 TCAAGCA 2255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2186 TAGAGCA 2192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 7

```

US-08-902-632-3
; Sequence 3, Application US/08902632
; Patent No. 608025
; GENERAL INFORMATION:
; APPLICANT: KOMATSUBARA, Hideyuki
; APPLICANT: KITABAYASHI, Masao
; APPLICANT: KAWAMURA, Hideki
; APPLICANT: KAWAKAMI, Bunsei
; APPLICANT: KAWAMURA, Yoshihisa
; APPLICANT: TAKAGI, Masahiro
; APPLICANT: IMANAKA, Tadayuki
; TITLE OF INVENTION: Modified Thermostable DNA Polymerase,
; AND DNA Polymerase Composition for Nucleic Acid
; TITLE OF INVENTION: Amplification
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1 Broadway
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.50 inch floppy disk, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/902,632
; FILING DATE: Concurrent Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 198911/96
; FILING DATE: 29-JUL-1996
; APPLICATION NUMBER: JP 200446/96
; APPLICATION NUMBER: 30-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: GREASON, Edward W.
; REGISTRATION NUMBER: 18,918
; REFERENCE/DOCKET NUMBER: 2418/7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-425-7200

```

```

; TELEFAX: 212-425-5288
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-902-632-3

```

```

Query Match 7.9%; Score 191.4; DB 3; Length 2325;
Best Local Similarity 48.0%; Pred. No. 3.4e-50;
Matches 854; Conservative 0; Mismatches 856; Indels 69; Gaps 8;

QY 622 GTGAAGATGACGATGGCGATGAGTGCTATTTCATTGTGAGAGGGGCAAGACGATCGAAAA 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 508 GTGATAACTTTGGAAGAAGCTGGATCTCCCTCCTACGTTGACGTGCTGCGACGAGAGGAG 567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 682 CGATACCGCGAGTTTGTAGAGTACCTGAAGAGGTATGACCCCGACATATAGTCGGTTAT 741
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 568 ATGATTAAGCGCTTCTCGGTGTGTGAAGGAGAAAGACCCGAGCTTCTCATACCTAC 627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 742 AACAAATCAATTCGATTTGGCTTATCTTTTGGAGGCGCGCCGCTAGGATGAAG 801
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 628 AACGGCGACAACTTCGACTTCGCTTCTGAAAAGCGCTGTGAAAAGCTCGGATTAAC 687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 802 CTGTATGTACTAGAAAGTTGGCGCGAGCCGACCTACCTGCTACATGGGACGCTCT 861
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 688 TTGCGCTCGGAAGGATGGAAGCGGAGGATTCAGAGGATTTGGGCGACAGGTTTGGC 747
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 862 GTCCCTGGCAGGCTTAAAGTAGATCTGTACGACTATGCGGAAGAGATGCCAGAGATCAAG 921
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 748 GTCGAAGTGAAGGAGCGGATACACTTCGATCTCTATCTGTGATGAAGCGACGATTAAC 807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 922 ATAAAGATCTCGAGGAGTCCAGAGTATCTAGGCGGTGTGAAGAGAGTGAACGGT 981
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 808 CTGCCACATACACGCTTGAGGCGCTTTATGAAGCGCTTTCGGTCAGCCGAA--GGAG 864
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 982 ATCATCAATTTGGTGGAGATTCAGACTATTTGGGCGACCCCGAGAGAGAGCCACTATTA 1041
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 865 AAGTTTACGTGAGGAAATTAACACAGCTTGGGAAACCGGCGAGAACTTTGAGAGAGTC 924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1042 CTGCAATACGCGCGGACGATGTCGCGCTACTTACGCGTTAGCCGAGAGATATTCGCG 1101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 925 GCGCGTACTCGATGGAAGTGCAGAGTCAATACAGAGCTTGGAGAGAGTTCCTTCG 984
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1102 TTGTATCCAGTTCTGTCGTAACAGTCTCCACCTAGACACGAGTGGTGGATGAT 1161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 985 ATGAGGCGCCAGCTTTCGCTTAAATCGGCGAGTCTCTCTGCGAGCTCCCGCTCCAGC 1044
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1162 GTTGCTTTTCGACTTGAATGTTACCTGATACGCGCGGCTTTAAGATGAAAGAGCTTGTG 1221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1045 ACTGGCAACCTCGTTGAGTGTCTCTCTCAGGAGGCTTATGAGAGGATGAGCTGCC 1104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1222 CGAACCCTGAGC-----GCCAGAGAGACTTACCGTGGCGCTATA 1266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1105 CCGAACACCCCGGATGAAAGAGCTGGCCGAGAGAGCGCAGAGCTATGAAGAGGCTAT 1164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1267 GTTCTTGAGCCCTTGAGAGGCGTGCACGAGATATAGCCGTACTCTGACTTTAGCTCGATG 1326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1165 GTAAGAGCCCGAGAGAGGTTGTGGAGAGAAATAGTGTACCTAGATTTAGATTCCTG 1224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1327 TACCCAAATCATGATAAAGTACAAATTTGGTCTGTGACAGCTTGTGAGCCCTGGTGA 1386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1225 TACCCCTCAATCATCATCACCCACAACTCTCGCGGATACGCTCAACAGAGAGGATGC 1284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1387 AGTGTGGCGAGTGTGTTGCTGGAGGCGCGCGGAGTCAAGCAAGTTCGATAGTGT 1446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1285 AAGGAATATGACGTT-----GCCCGACAGGTGCGCCACCGCTTCTGCAAGGAC 1332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1447 CCGCCCGGCTTCTCAAGACAGTCTCTGAGAGGCTGTTAGAGCTTCGTAACGCTGTCGT 1506
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1333 TTCCAGGATTTATCCGAGGCTGTTGGAGACCTCTTAGAGAGGAGGAGGAGCAAA--GATA 1389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

QY 1507 GCTGAATGAGAGATATCTCCGATAGCCAGATATCGACTGTTGGATGAAGCAG 1566
Db 1390 AAGAGAGAGATGAAGCCACGATGACCCGATCGAGAGAACTCTCGATTACAGGAG 1449
QY 1567 AAGCGTTGAAGGTTCTTTCGAAAGCTAGTTACGGCTACATGGGTTGGAGCGCGCTAGG 1626
Db 1450 AGGCGCATAGATCTCGCAACACAGCTACTACGGTTACTACGGCTATGCAAGGCGCGC 1509
QY 1627 TGGTATTGAGGAGTGGCGAAAGCTGTACGGCTTGGGTAGGACCTCATAGCACCC 1686
Db 1510 TGGTACTGCAAGGAGTGTGAGAGAGCGTAAACGGCTGGGAGGAGTACATACCATG 1569
QY 1687 GCGATCA--ACATAGCTCGTAAACTAGGCTCAAGGTGATCTAGGTTGACACAGATTGG 1743
Db 1570 ACCATCAAGGAGATAGAGGAAGTACGGCTTTAGGTAATCTACAGGACACCGACGA 1629
QY 1744 CTCTCGTGACCTATGATCCGAGAGAGGTGAAATTTTCATCAAAATTTAAAGGAGAG 1803
Db 1630 TTTTTCGACATACCTGGAGCGGATGCTGAACCGTCAAAAGAGGCTATGGAGTTC 1689
QY 1804 CTG-----GGTTGGAATCAAGCTAGAGAGGTGTAC 1836
Db 1690 CTCACATATATCAAGCCCAAACTTCCGGCGCGCTTGAGCTCGAGTACGAGGGCTTCTAC 1749
QY 1837 AAACGCTTATCTTTACAGAGGCTAAGAGAGGTACGCTGGCTTCTCGAGGACGGAGCT 1896
Db 1750 AAACGCGCTTCTTC---GTGCGAGAGAGATGCGGTGATAGACGAGAGGACG 1806
QY 1897 APAGATATGTCGGTTTCGAGCTGTACGTGCGGATGTTGTTGAACTCGCCCAAGAGGTT 1956
Db 1807 ATAAACAACGCGGAGCTTGTAGATTGTGAGGCGTGTCTGAGCGAGATAGCGAAAGAGACG 1866
QY 1957 CAGACTAGGTTGTGGAATAGTATTGAAGAGGAGTGTGAGTGTGACAGGCTGTAGACTAC 2016
Db 1867 CAGCGAGGGTTCTTGAAGCTTTGCTAAAGGACCGGTGACGTCGAGAGGCGCTGAGGATA 1926
QY 2017 GTCAGAGATGTTGAAGAGTTGAGGAGGCGAGGTTTCCCATAGAGAACTTGTATC 2076
Db 1927 GTCAGAGATGTTACGGAAGTGTAGCAAGTGTAGAGGTTTCCCGGAGAGACTGTTGATC 1986
QY 2077 TGGAGAGCCCTTAGTAAGGCTTCTGAGGAGTACACAAACGAGGACCAACACGCTGTGCA 2136
Db 1987 CACGAGCAGATAACGAGGATTTAAAGGACTACAGAGGCAACCGGCTCCCAAGTTGCCGTT 2046
QY 2137 GCGAGAGAGTCTGTACGAGGCTACCGGTAAGCCAGCGAGGACCAACGCTGTAGTGTGTA 2196
Db 2047 GCGAGAGGTTGGCCGCGAGAGGAGTCAAAATACCGCTTGAACCGGTGATAGCTACATC 2106
QY 2197 ATAGTGAAGGTTGGCGGTATCAGTCAAGAGCATGGCCATCTTCATGCTCAAGGAT 2256
Db 2107 GTGCTCAAGGCTCTGGGAGGATAGCGACAGGCGGATACCGTTGACGAGTTGACCCG 2166
QY 2257 CCGACCGAGATAGAGTGACC---TACTATGTTGACCAACCAATCATCTCCGGCTGCAATG 2313
Db 2167 ACAGAGCAGACAGTACGACGCGGAGTACTACATTTGAGAAACGAGTTTCTCCGAGCGGTTGAG 2226
QY 2314 AGAATATGGGCTACTTTGGCATCACCGGAGAGAGCTG 2352
Db 2227 AGAATCTGAGAGCTTCTGGTTACCGCAAGGAGACCTG 2265

RESULT 8

US-08-906-925-5

; Sequence 5, Application US/08906925

; Patent No. 5882904

; GENERAL INFORMATION:

; APPLICANT: Riedl, William

; APPLICANT: Fly, Susan

; APPLICANT: Kaborod, Barbara F.

; APPLICANT: Nye, Steven H.

; APPLICANT: Ting, Eve

; TITLE OF INVENTION: THERMOCOCCUS BAROSII DNA POLYMERASE MUTANTS

NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: U.S.A.
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,925
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Baker, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 740211.90628
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5709
TELEFAX: (414) 271-3552
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2337 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-906-925-5

Query Match 6.7%; Score 162.2; DB 2; Length 2337;

Best Local Similarity 45.7%; Pred. No. 6.5e-41;

Matches 894; Conservative 0; Mismatches 998; Indels 63; Gaps 7;

QY 443 ACAAGATGGCTTCCGTGTCGACAAAGGTATACCTGCTTAAGACAGCGCGGAGCCACTTT 502
Db 329 ACAGTAGACATACCTTTCGCCAAGCGGTACTCTATGACAGAGGTCTCGTCCCGATGG 398
QY 503 ATGCTGAGGCTCTCGACCAACCAAGCTTCCCGATCTTAGGATATCTCGGTTTCGATATTG 562
Db 389 AGGCGCATGAGGAGCTTAAACTCATGCTCTGACATCGAGACGCTCTACACGAGGAG 448
QY 563 AGCTTTATAGCAAGCAAGGTCGCCGCTCCAGAGCGGATCTCTGTAATAGTATAGCTG 622
Db 449 AAGAGTTCGGAACCGGCGCGATTTCTGATGATAGCTACGCGATGAAGAGCGGCGCGTG 508
QY 623 TGAAGACTGACGATGGCGATGAGTGTCTATTTCATTCGAGAGGCGCAAGACGATCGAAAC 682
Db 509 TGATACTCTGGAAGAGATCGACCTGCCCTACGTCGACGTTGTCCTCCACGAGAAGAGA 568
QY 683 CGATACGCGAGTTTGTAGATGATGTAAGAGGATGACACCCGACATAAATAGTCGGTTATA 742
Db 569 TGATAAGCGCTTCTTAAAGTCTGTTAAGGAGAAGGACCCGAGCGCTGCTGATTAACATACA 628
QY 743 ACAACATCATTTTCGATTGGCTTATCTTTTGAAGGCGCGCCCGCATCTTAGGCATAAAGC 802
Db 629 ACGCGCAACTTTCGACTTCGCTTACTCTAAAGCGGTTGTGAGAAGCTTGGCGTGAGCT 688
QY 803 TTGATGTGACTAGAAGAGTTGGCGCGAGCCCACTAGCTGATGATGGCAGCTCTCTG 862
Db 689 TTACCTTCGCGAGGAGCGGAGCGAGCGCAAGATACAGCGCATGGGCGACCGCTTCGCGG 748
QY 863 TCCCTGCGAGGCTTAACGTATGATCTGTACGACTATGCGGAAGAGATGCCAGAGATCAAGA 922
Db 749 TTGAGGTGAAGGCGAGGATCCACTTCGACCTGTACCCCTCATAGGCGCAGCCATAAAC 808
QY 923 TAAAGAGTCTCGAGGAGGTCGAGAGTATCTAGGCGGTGATGAAGAGAGGTGACCGGTTA 982
Db 809 TCCGACCTACACCTTGAAGGCTGTATACAGGCGGTTTTCGCAAGGCCAA---GGAGA 865

QY	983	TCATCAATTGTTGGGAGATTCCAGACTATTATGGGACGACCCGAAAGAGAGACCACTATTAC	1042
Db	866	AGGTCTACGCCGAGAGATACCCACCGCTTCGGAGACCGCGAGGGGCTTTGAGAGGTCG	925
QY	1043	TGCAATACGCGCGACGANTGTCGCGCTACTTACGCTTAGCCGAGAAATATTGCCGT	1102
Db	926	GCGCTACTCATGAGACGCGAGGGTTACTTACGAGCTTGGCAGGGAGTTCTTCCCGA	985
QY	1103	TTGCTATCCAGTTGTCGTACGTTACAGGCTCTCCACATAGACCAGGTAGGTGCGATGAGTG	1162
Db	986	TGAGGGCCACGCTTTCAGGCTCATCGCCAGGGTCTCTGGAGCTTTCGCCCTCAGCA	1045
QY	1163	TTGCTTTGACTTGAATGTTACTGATACGCGCGCGTTTAAGATGAAGAAGCTTTGTGC	1222
Db	1046	CCGCAACCTTGTGTGATGTTTGTCTCAGAAAGCTTACGAGAGACGAACCTCGCTC	1105
QY	1223	CGAAC-----CGCGTTGAGCGCCGAGAGACTTACCGTGGCGCTATAG	1267
Db	1106	CCACAAGCCGACGAGAGGAGCTGCGAGGAGAGAGGGGGGCTACGCCGTGGCTACG	1165
QY	1268	TTCTTGAGCCGTTGAGAGGGGTGCACGAGAAATAGCCGTACTCGACTTTAGCTCGATGT	1327
Db	1166	TCAAGGACCGGACGGGACTGTGGATATATTGTGTACTCTCGATTTTCGCTCGCTGT	1225
QY	1328	ACCCAAACATCATGAATAAGTACATGTTGGTCTCGACGCTGTGAGCCCTGGTGA	1387
Db	1226	ACCCCTCCATTATCATCCCAACAGCTCTCGCAGATACGCTCAACCGCGAGGATGTA	1285
QY	1388	AGTGTGGGAGTGTGTTGTTGGGAGGCCCGGAGGTCAAGCACAGGTTCCGTAGGTGTC	1447
Db	1286	AGACTACGAGCTT-----GCCCGCAGGTCGCTCAAGTTCTCGAAGACT	1333
QY	1448	CGCCCGCTCTTTCAGACAGTTCCTTGAGAGGCTGTTAGAGCTTCGTAAGCGTGTGCTG	1507
Db	1334	TCCCGGCTTCATTCGAGCCTGCTCGAAACCTGCTGGAGGAGAGCGAAGATTAAGA	1393
QY	1508	CTGAATGAAGAAGTATCTCTCGGATAGCCGAATATCGACTGTGTGATGAAGGCAGA	1567
Db	1394	GGAAGATGAAGGC---AAGCTCGACCGCTGGAGGAAAGCTTCTCGATTATCGCCAGC	1450
QY	1568	AGCGTTGAGGTTCTTGCAAAACGCTAGTTACCGCTTACATGGTGTGAGCGCGCTAGGT	1627
Db	1451	GCGCTATCAAAAATCTCTGGCGAAGCTTCTACCGCTATTACCGCTACGCCAGGCGCAAGAT	1510
QY	1628	GGTATTACGAGGAGTGCAGAAAGGCTGTACGCTTGGGTTAGGCACTTATACGACCG	1687
Db	1511	GGTACTGCAAGAGTGCAGAGCGTTACGCGATGGGGCAGGGATCATCGAATGG	1570
QY	1688	CCATCA---ACATAGCTCGTAAACTAGGCTCAAGGTGATCTACGGGTGACACAGATTCCG	1744
Db	1571	TTATCAGAGAGCTTGAGGAAAAGTTGCGTTTAAAGTCTCTATGCAGACACAGACGGTC	1630
QY	1745	TCITTCGTGACCTATGATCCGGAAGGTGGAATAATTATCAAAAATTTAAGAGGGA--	1802
Db	1631	TTCAATGCCCAATTCCTTGGAGCGGACGCTGAAACAGTCAAGAAAAAGGCAATGGAGTTCT	1690
QY	1803	-----GCTCGGGTTTCAAANTCAAGCTTAGAAGAAGGTCTACAAC	1840
Db	1691	TAAACTATATCAATCCCAACTCCCGGCTTCTCGAACTCGAATACGAGGGCTTCTACG	1750
QY	1841	GCTTATCTTTTACAGAGGCTTAAGAAGGTACGCTCGGCCCTTCTCGAGGACGACGCTATAG	1900
Db	1751	TCAGGGCTTCTTCGTCAAGAAAGATACGCGTTATAGACGAGGAGGGCAAGATAA	1810
QY	1901	ATATTGTCGGTTTCGAGGCTGTACGTGGCGATTGGTGTGAACCTCGCCAAAGGAGTTTCA	1960
Db	1811	CCAACGCGCGGGCTTGAGATAGTTAGGAGGACTTGGAGCGAGATAGCGAAGGAGACG	1870
QY	1961	CTAAGGTGTGCGAAATAGTATTCGAACGAGGTGAGGTGAACAGGCTGTAGTAGTACGTCA	2020
Db	1871	CGAGGGTTCTTGAGGCGATACCTCAGGACGGTGACGTTGAGGAGGGCCGTTCAGANT	1930
QY	2021	GGAAGATGTGAAAGAGTTGGAGGAGCGCAAGGTTCCCATAGAGAAGCTTGTAAATCTGGA	2080

Db	1931	AGGAAGTGCAGCGAAAGCTGACGAAGTACGAGGTTCCGCGGAGAGCTGGTTATCCACG	1990
Qy	2081	AGACCCCTTAGTAGCGCTCTTAGAGAGTACACACGAGGACACACAGCTCGTTGCGACGGA	2140
Db	1991	AGCAGATATACGCGGAGCTCAAAAGCTTACAAAGCCACCGGCCCGCGCATGCGCATAGCGA	2050
Qy	2141	AGAGGATGCTGTCAGCAGGCTACCGGGTAAGCCAGCGCAAGATAGGTATGTTAATAG	2200
Db	2051	AGCGCTCGCGGAGGGGAATTAAGATACGCCCCCGGACGGTGAATAGTACATCGTCC	2110
Qy	2201	TGAAGGTTGTTGGCGGTATCAGTCAAGAGCATGGGCCATCTTCATGGTCAAGGATCTTA	2260
Db	2111	TCAAGGGCTCGGGGAGAATAGGCGACAGGGGCCATTCCCTTCGACGAGTTCGATCCGACGA	2170
Qy	2261	GCCA---GATGACGTGACCTACTATGTTTGACACCAAAATCATCCCGGCTGCATTGAGAA	2317
Db	2171	AGCACAGGTACGACGTGACTACTACTCATCGAAGACCGAGTTCTTCCAGCGGTGGAGAGAA	2230
Qy	2318	TACTGGGCTACTTTGGCATCACCGAGAAGCTG	2352
Db	2231	TCCTCAGGGCTTTGGCTACAGAGGAGACCTG	2265

RESULT 9

US-08-375-134-13
Sequence 13, Application US/08375134
Patent No. 5602011
GENERAL INFORMATION:
APPLICANT: Luhm, Robert A.
APPLICANT: d'Hennezel, Olga B.
APPLICANT: Duffaud, Guy D.
APPLICANT: Jolly, James F.
APPLICANT: Kelly, Robert M.
APPLICANT: King, Eve Yi-Fay
TITLE OF INVENTION: PURIFIED THERMOCOCCUS BAROSSII DNAPOLYMERASE
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: U.S.A.
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/375,134
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Baker, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 740211.90440
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5709
TELEFAX: (414) 271-3552
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 2734 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-375-134-13

7:

ADDRESSEE: Quarles & Brady
 STREET: 411 East Wisconsin Avenue
 CITY: Milwaukee
 STATE: Wisconsin
 COUNTRY: U.S.A.
 ZIP: 53202-4497
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/15263
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Baker, Jean C.
 REGISTRATION NUMBER: 35,433
 REFERENCE/DOCKET NUMBER: 740211.90440
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (414) 277-5709
 TELEFAX: (414) 271-3852
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2734 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 PCT-US95-15263-13

Query Match 6.1%; Score 146.2; DB 5; Length 2734;
 Best Local Similarity 45.2%; Pred. No. 8.7e-36;
 Matches 884; Conservative 0; Mismatches 1008; Indels 63; Gaps 7;

QY	443	ACNAGATGGCTCCGTCGACAAAGGTTCCCGATCTTAGGATCTCGGTCGATATTG	502
DB	565	ACGATACGACATACCTTTGCCAAGCGATATCTCATGACAAAGGCTTGATTCCTATGG	624
QY	503	ATGGTGAGGCTCTCGCACCAACCAAGCTTCCCGATCTTAGGATCTCGGTCGATATTG	562
DB	625	AGGCGGATGAGAGCTTAACATCTATGCTTCGATGATGATGATGATGATGATGATG	684
QY	563	AAATTTATAGCAAGCAAGGTCGCGCGCTCCAGAGCGGATCTCTGTAATAGTATGATG	622
DB	685	AAGAGTTCGAAACCGGCGGATCTGATGATGATGATGATGATGATGATGATGATG	744
QY	623	TGAACATGACATGGCGATGAGTGCTATTCTATTCGAGAGGCAAGCAAGCATCGAAAC	682
DB	745	TGATTAACCTGGAAAGATCGACCTCGCCCTAGCTGACGTTCTCTCCACCGAAGGAGA	804
QY	683	CGATACGGGAGTTTGTAGATGATGTAAGAGTATGACCCGACATATAGTCGGTTATA	742
DB	805	TGATAAAGCGCTTCTTAAGTCTGTAAGGAGAGGACCGGACGCTGATGATGATGATG	864
QY	743	ACAAATATCTTCAATGGCTTATCTTTTGAAGCGCGCGCATCTTAGGATGATGATG	802
DB	865	ACGGGCAACACTTCGACTTCGCTTACCTCAAAAGCGGTGTGAGAGCTTGGCGTGA	924
QY	803	TTGATGCTACTAGAGAGTTGGCGGAGCCCACTAGGCTGATGAGGCAAGTCTGCTG	862
DB	925	TACCTCGGAGGAGCGGAGAGGACCGGACGATGATGATGATGATGATGATGATGATG	984
QY	863	TCCCTGGCAGGCTTAACGTAGATCTGTACGACTATGCGGAGAGATGCGAGATCAGA	922
DB	985	TTGAGTGAAGGCGAGGATCCACTTCGACCTGTACCCCGCATAGGCGCACCATTAACC	1044
QY	923	TAAAGATCTCGAGAGGTCGACAGTATCTAGGCGTGTATGAGGAGAGTGAACCGCTTA	982
DB	1045	TCCCGACCTACACCTTGAGGCTGTATACGAGGCGGTTTTCGGCAGCCCA---GGAGA	1101
QY	983	TCATCAATGGTGGAGATTCAGACTATTGGGACGACCCCGAAGAGACCATATTATAC	1042

DB	1102	AGGTCTACGCCGAGGAGATAGCCACCGCTTGGGAGACCGGTGAGGGGCTTGAGGGGTCG	1161
QY	1043	TGCAATACGCGCGGACGATGTCCGCGCTACTTACGCTTAGCCGAGAGATATTGCCGT	1102
DB	1162	CGCGTACTCGATGGAGGACGCGAGGGTTACTACGAGCTTGGCAGGAGTTCTTCCGA	1221
QY	1103	TGCTATCCAGTTGTCGTACCTAACAGCTCTCCACTAGACACAGTAGTCGATGATG	1162
DB	1222	TGAGGCCAGCTTTCCAGGCTCATCGCCAGGGTCTTGGGAGCTTTCCGCTCCAGCA	1281
QY	1163	TGCTTTTCGACTTGAATGTGCTGATACCGCGGGCTTTAAGATGAAGAGCTTGTGC	1222
DB	1282	CGGCAACCTTGTGAGTGGTTTGTCTCAGAAAGCTACGAGGAGCACTCGCTC	1341
QY	1223	CGAAC-----CGCTTGAGCCGCCAGAGAGACTTACCGTGGCGCTATAG	1267
DB	1342	CCAACAAGCCCGACGAGGAGGAGCTGCGAGGAGAGAGGGGGGCTACCGCGTGGCTACG	1401
QY	1268	TTCTTGAGCGCTTGAGAGGCGTGCACGAGATATATCCGCTACTCGACTTTAGCTCGATG	1327
DB	1402	TCAAGGAGCCGAGCGGGGACTGTGGATATATTTGTACCTCGATTTTTCGCTCGCTGT	1461
QY	1328	ACCAAAATCATGATTAAGTACAAATGTGTCTCTGACAGCTTGTGAGGCTTGTGAAA	1387
DB	1462	ACCCCTCCATTATCATCACCAACAGCTCTCGCCAGATACGCTCAACCGGAGGAGTGA	1521
QY	1398	AGTGTGGGAGTGTGTTGCTGGAGGCGCGGAGGTCAGACAGCTTCCGTAGTGTGTC	1447
DB	1522	AGGCTACGAGCTT-----GCCCGCAGGTCGCTCACAAGTTTCTGCAAGACT	1569
QY	1448	CGCCCGCTTCTTCAAGACAGTTCTTTCGAGAGGCTTTAGAGCTTCTGTAAGCTGTGGTG	1507
DB	1570	TCCCGGCTTCATTCGAGGCTGCTCGAAGAACCTCTGAGGAGAGAGGCAAGATAAGA	1629
QY	1508	CTGAATTAAGAGATATCTTCCGATAGCCAGAAATATCACTGTTTGGATGAAGGAGA	1567
DB	1630	GGAGATGAAGGC---AACGCTGACCCGCTGGAGAGAGCTTCTGATTTCCGAGC	1586
QY	1568	AGGCTTGAAGTTCTTTCGAAAACGCTAGTTACGGCTACATGGGTGGAGCGGCTAGGT	1627
DB	1687	GCCTATCAAAATCTCGCGAAGAGCTTCTACGCTATTACGCTACGCCAGGAGAT	1746
QY	1628	GGTATTCAGGAGTGCAGAAAGCTGTCAGCGCTTGGGTGAGGAGGCTTATAGGACCG	1687
DB	1747	GATCTCAAGAGAGTGGCGAGAGCTTACGGGATGGGCGAGGAGTATACATAATGG	1806
QY	1688	CCATCA---ACATAGCTCTGTAACACTAGCGCTCAAGGTGATCTACGGTGACACAGATTGC	1744
DB	1807	TTATCAGAGAGCTTGAGGAAAGTTCGGTTTAAAGACCTCTATGAGACACAGACGGTC	1866
QY	1745	TCTTCGTGACTATGATCCGAGAGAGTGGAAATTTTCAATAAATTAATAAGGAGA--	1802
DB	1867	TTATGCCACCATCTCTCGAGCGGACCGGTGAACAGCTCAAGAAAGGACCTGAGCTTCT	1926
QY	1803	-----CCTCGGGTTCGAAATCAAGCTTCTCGAACTCGAATACGAGGCTTCTACT	1986
DB	1927	TAACTATATCAATCCCAACTCCCGGCTTCTCGAACTCGAATACGAGGCTTCTACT	1986
QY	1841	GCTTATCTTTACAGAGGCTTAAGAGAGTACGCTGCGCTTCTCGAGGAGGAGCTATAG	1900
DB	1987	CGAGGGCTTCTTCGTCCACCAAGAGAGTACGCGCTTATAGACGAGGAGGAGGAGATAA	2046
QY	1901	ATATTCTCGTTTCGAGGCTGATGAGGAGTGTGGTGTGTAATTCGCAAGGAGGTTTCA	1960
DB	2047	CCACGCGCGGCTTGAGATAGTTAGGAGGAGCTGAGGAGATAGCGAAGGAGAGCTG	2106
QY	1961	CTAAGGTTGCGAAATAGTATTCAAGCAGGTCAGTGAACAGGCTTAGAGTATGCTCA	2020
DB	2107	CGAGGGTCTTTCGAGGCTTACTCAGGACCGGTGAGCTTGGAGGCGCTCAGAACTGCTCA	2166
QY	2021	GGAGATTTGTAAGAGTTGGAGGAGGAGGAGGCTTCCCATAGAGAGCTTGTAACTGGA	2080
DB	2167	AGGAAGAGACGGAAGAGCTGAGCAAGTACGAGGTTCCCGCGGAGAGCTGTTTATCACCG	2226

QY 2081 AGACCTTAGTAAAGCGTCTTGAGGAGTACACAAAGGAGGACACACACGTCCTTGCAGCA 2140
|||
DB 2227 AGCAGATAACCGCGAGCTCAAGAGCTACAAGGCGACCGCGCGACGTCGCGCATAGCGA 2286
|||
QY 2141 AGAGGATGCTGTCAGCAGGCTACCGGGTAAGCCAGGCGACAAAGATAGGGTATGTAATAG 2200
|||
DB 2287 AGCGCTCGCGCGAGGGAATAAGATACGCGCGGACGCGTATAGCTACATCGTCC 2346
|||
QY 2201 TGAAGGTGGTGGCGGTATGATCAAGAGATAGGCGCATCTTCAATGATCAAGATCCTA 2260
|||
DB 2347 TCNAGGCTCGGGGAGATAGCGACAGGCGCATTCCTTCGACGAGTTCGATCCGACGA 2406
|||
QY 2261 GCCA---CATAGACGTACCTACTATGTGTGACCAACCAATCATCCGCGCTCATAGAA 2317
|||
DB 2407 AGCAGAGTACGACGCTGACTACTACATCGAGAACCAAGTTCCTCCAGCGGTGGAGAGAA 2466
|||
QY 2318 TACTGGGCTACTTTGGCATCACCGGAGAGAGCTG 2352
|||
DB 2467 TCCTCAGGCGCTTCGGCTACAGAGAGAGACCTG 2501
|||

RESULT 11

US-09-073-354-6
; Sequence 6, Application US/09073354
; Patent No. 6033859

GENERAL INFORMATION:

; APPLICANT: KITABAYASHI, Masao
; APPLICANT: ARAKAWA, Taku
; APPLICANT: INOUE, Hiroaki
; APPLICANT: KAWAKAMI, Bunsei
; APPLICANT: KAWAMURA, Yoshihisa
; APPLICANT: IMANAKA, Tadayuki
; APPLICANT: TAKAGI, Masahiro
; APPLICANT: MORIKAWA, Masaaki
; TITLE OF INVENTION: A Thermostable DNA Polymerase and Kits for
; TITLE OF INVENTION: Amplifying Nucleic Acids
; NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

; ADDRESSES: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036

COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1 Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,354
; FILING DATE: Concurrent Herewith
; CLASSIFICATION: 0506

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/656,005
; FILING DATE: 24 MAY 1996
; APPLICATION NUMBER: JP 334096/95
; FILING DATE: 31 MAY 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Toffenetti, Judith L.
; REGISTRATION NUMBER: 39,048
; REFERENCE/DOCKET NUMBER: 2418/9

TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-429-1776
; TELEFAX: 202-429-0796

INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 5339 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA

ORIGINAL SOURCE: Hyperthermophilic archaeon

US-09-073-354-6

Query Match 5.4%; Score 131.4; DB 3; Length 5339;
Best Local Similarity 52.2%; Pred. No. 6.7e-31;
Matches 412; Conservative 0; Mismatches 341; Indels 36; Gaps 4;

QY 1597 TAGCGGTACATGGGTTGGAGCGCGTAGTGGTATTGACGAGGATGCGCAAGGCTGTC 1656
|||
DB 4320 TAGCGGTACTACCGCTATGCAAGGGCGCGCTGCTACTGCAAGGAGTGTGCAGAGACGTA 4379
|||
QY 1657 ACCGCTTGGGTTAGGACCTCATACGACCGCCATCAAC---ATACCTCGTAAACTAGGC 1713
|||
DB 4380 ACCGCTTGGGTTAGGAGGTACATACGATGACCATCAGAGATAGAGAAAGTACGGC 4439
|||
QY 1714 CTCAGGTGATCTACCGGTGACACAGATTCGCTCTTCGTCACCTATGATCCGAGAGAGTG 1773
|||
DB 4440 TTTAAGGTATCTACAGCGACACCGACGGATTTTTTGCACATACTCTGGAGCGGTGCT 4499
|||
QY 1774 GGAATTTTCATCAAAATTAAGAGGAGGAGCTG----- 1806
|||
DB 4500 GAAACCGTCAAAAAGAGGCTATGAGATTCCTCACTATATCAAGCCCAAACTTCGCGGC 4559
|||
QY 1807 GCGTTGAAATCAAGCTAGAGAGGTGTACAAACGCTTATCTTTACAGAGGCTAAGAG 1866
|||
DB 4560 GCGTTGAGCTCGAGTACGAGGCGCTTCTACAAACGCGGCTTCT---CGTCAGAAAG 4616
|||
QY 1867 AGGTAGCGTGGCCTTCTCGAGGAGCGACGTATAGATATTGTGCGTTTCGAGGCTGTACCT 1926
|||
DB 4617 AAGTATCGGTGATAGACGAGGAGCAAGATAACAACCGCGGACTTGAGATTGTGAGG 4676
|||
QY 1927 GCGGATTTGGTGAATCTGCCAAGGAGGTTACAGCTAAGGTTGTCGAAATAGTATTGAAG 1986
|||
DB 4677 CGTGACTGGAGCGAGATAGCGAAAGAGACGCGAGCGGAGGTTCTTGAAGCTTTGCTAAA 4736
|||
QY 1987 ACCAGTGAAGTGAACAAGGCTGTAGAGTAGCTCAGCAAGATGTGAAAGTTGGAGGAG 2046
|||
DB 4737 GACGTGACGTGAGAGAGGCGGTGAGGATAGTCAAAGAGTTACCGAAAGCTGACCAAG 4796
|||
QY 2047 GGCAAGGTTCCCATAGAGAGCTTGAATCTGGAAGACCTTAGTAAGCGTCTTGAGGAG 2106
|||
DB 4797 TAGAGGTTCCGCGGAGAGCTGGTGATCCACGACGACATACGAGGAGTTTAAAGGAC 4856
|||
QY 2107 TACACACGAGGACACACACGCTCTGCGAGCAAGAGATGCTGTACAGAGCTACCGG 2166
|||
DB 4857 TACAAGGCAACCGGTTCCCGCTGCGGTTGCCAAGAGGTTGCCCGAGAGGAGTCAA 4916
|||
QY 2167 GTAAAGCCAGCGCACAAAGATAGGCTATGTAATAGTGAAGGTTGGTGGCGGTATCACTCA 2226
|||
DB 4917 ATACGCCCTGGAACGGTGATAAGCTACATCGTCTCAGGGCTCTCGGAGGATAGCGGAC 4976
|||
QY 2227 AGAGCATGGCCATACCTTCATGTCAGAGGATCCTAGCAGATAGACGCTGACC---TACTAT 2283
|||
DB 4977 AGGCGGATACCGTTTCGACGAGTTTCGACCGGACGACCAAGTACGATGCGGAGTACTAC 5036
|||
QY 2284 GTTGACCAACCAATCATCCCGCTGCATTGAGATATCTGGCTACTTTCGCATCACCAG 2343
|||
DB 5037 ATTGAGAACCCAGGTTCTCCCGCGCTTGAGAGAAATTCGAGAGCCTTCGTTACCGCAAG 5096
|||
QY 2344 AAGAAAGCTG 2352
|||
DB 5097 GAAGACCTG 5105
|||

RESULT 12

US-08-656-005A-6
; Sequence 6, Application US/08656005A
; Patent No. 6054301

GENERAL INFORMATION:

; APPLICANT: KITABAYASHI, Masao
; APPLICANT: ARAKAWA, Taku
; APPLICANT: INOUE, Hiroaki
; APPLICANT: KAWAKAMI, Bunsei

APPLICANT: KAWAMURA, Yoshihisa
 APPLICANT: IMANAKA, Tadayuki
 APPLICANT: TAKAGI, Masahiro
 APPLICANT: MORIKAWA, Masaaki
 TITLE OF INVENTION: A Method of Amplifying Nucleic
 TITLE OF INVENTION: Acid and A Reagent Therefor
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Kenyon & Kenyon
 STREET: 1025 Connecticut Avenue, N.W., Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WordPerfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/656,005A
 FILING DATE: 24 MAY 1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 134096/95
 FILING DATE: 31 MAY 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Toffenetti, Judith L.
 REGISTRATION NUMBER: 39,048
 REFERENCE/DOCKET NUMBER: 2418/3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-429-1776
 TELEFAX: 202-429-0796
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5339 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE: Hyperthermophilic archaeon
 US-08-656-005A-6

Query Match 5.4%; Score 131.4; DB 3; Length 5339;
 Best Local Similarity 52.2%; Pred. No. 6.7e-31;
 Matches 412; Conservative 0; Mismatches 341; Indels 36; Gaps 4;

QY	1597	TACGGCTACATGGTTGAGCGCGCTAGGTGGTATTGACGGAGTGCCTGCTC	1656
Db	4320	TACGGTTACTACGGCTATGCAAGGCGCGCTGTCTGCAAGGAGTGTGCAGAGAGCGTA	4379
QY	1657	ACGGCTTGGGGTAGGCACCTCATACGACCGCCATCAC---ATAGCTCGTAAGCTAGGC	1713
Db	4380	ACGGCTTGGGAGGAGGTACATAACGATGACCATCAAGGAGATAGAGAAAGTACGGC	4439
QY	1714	CTCAGGTGATCTACGGTGACACAGATTGGCTCTTCTGACCTATGATCCGGAGAGGTG	1773
Db	4440	TTTAAGGTAATCTACAGCGACACCGACGGAATTTTGGCAATACCTGGAGCGATGCT	4499
QY	1774	GAAATTTATCAAAATTAAGGAGGAGCTG-----	1806
Db	4500	GAAACCGTCAAAAGAGGCTATGAGTTCTCAACTATATCAACGCCAACTTCCGGGC	4559
QY	1807	GGTTTCGAAATCAAGCTAGAGAGGTGTACAAAGCTTATTTCTACAGAGCTTAAGAG	1866
Db	4560	CGCTTGAGCTCGAGTACGAGGGCTTCTACAAACGCGCTTCTT---CGTCACGAAGAG	4616
QY	1867	AGGTACGCTGGCTTCTCAGAGACGAGCTATAGATATTGTCGGTTTCGAGCTGTAGT	1926
Db	4617	AAGTATCGGTGATAGACGAGGAGCGAGATACACCGCGGACTTGAATTTGAGG	4676
QY	1927	GCGATTTGGTGTGAATCTCGCAAGGAGGTTTCAGACTAAGTTGTGCAATAGTATTGAAG	1986

RESULT 13

US-09-073-259-6
 ; Sequence 6, Application US/09073259
 ; Patent No. 6143536
 ; GENERAL INFORMATION:
 ; APPLICANT: IMANAKA, Tadayuki
 ; APPLICANT: TAKAGI, Masahiro
 ; APPLICANT: MORIKAWA, Masaaki
 ; TITLE OF INVENTION: DNA Encoding A Thermostable DNA Polymerase
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Kenyon & Kenyon
 ; STREET: 1025 Connecticut Avenue, N.W., Suite 600
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WordPerfect 6.1 Windows
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/073,259
 ; FILING DATE: Concurrent Herewith
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/656,005
 ; FILING DATE: 24 MAY 1996
 ; APPLICATION NUMBER: JP 134096/95
 ; FILING DATE: 31 MAY 1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Toffenetti, Judith L.
 ; REGISTRATION NUMBER: 39,048
 ; REFERENCE/DOCKET NUMBER: 2418/10
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-429-1776
 ; TELEFAX: 202-429-0796
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5339 base pairs
 ; TYPE: nucleic acid

Db	4677	CGTGACTGGAGCGAGATAGCGAAAGAGACGACGCGAGGGTTCCTTGAAGCTTTGCTAAAG	4736
QY	1987	ACGAGTGGAGTGAACAAGGCTGTAGAGTACGTCAGGAAGATTGTGAAGAGTTGGAGGAG	2046
Db	4737	GACGCTGACGTCGAGAAGGCGGTGAGGATAGTCAAGAAGTTACCGAAAGCTGAGCAAG	4796
QY	2047	GGCAAGGTTCCCATAGAGAGAGCTTGTAACTCTGGAAGACCTTAGTAAGCGCTTTGAGGAG	2106
Db	4797	TACGAGGTTCCGCGGAGAAAGCTGGTGTATCCACAGCAGATAACGAGGATTTAAAGGAC	4856
QY	2107	TACACAAGCGGAGGACCAACACGCTGTTCCAGCGAGAGGAGTGTGTGACGAGGCTACCGG	2166
Db	4857	TACAAGCAACCGGTCCCGAGGTTGCCGCGAGAGGAGTGTGCGCGAGAGGAGTCAA	4916
QY	2167	GTAAGCCGAGCGACAGATAGGATGTATTAAGTGAAGGTTGGTGGCCGCTATCAGTCAA	2226
Db	4917	ATACGCGCTTGGAAACGCTGATAAGCTACATGCTCAAGGCTCTGCGGAGATAGGCGAC	4976
QY	2227	AGAGCATGGCCATATCTTCATGTCAGGTCAGGTCCTAGCCAGATAGACGTGACC---TACTAT	2283
Db	4977	AGGCGATACCGTTCCGAGGAGTTCGACCCGACGAGCAAGTACGATGCGGAGTACTAC	5036
QY	2284	GTTGACCAACCAATCATCCCGGCTGCATTGAGATATCTG33CTACTTTGGCATCACCGAG	2343
Db	5037	ATTGAGAACGAGTTCTCCAGCGCTTGAGAGATTTCTGAGAGATTTCTGAGAGCTTCGGGT	5096
QY	2344	AAGAAGCTG 2352	
Db	5097	GAAGACCTG 5105	

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE: Hyperthermophilic archaeon
US-09-073-259-6

Query Match 5.4%; Score 131.4; DB 3; Length 5339;
Best Local Similarity 52.2%; Pred. No. 6.7e-31;
Matches 412; Conservative 0; Mismatches 341; Indels 36; Gaps 4;
QY 1597 TACGGCTACATGCTGGAGCGCGCTAGGTGGTATTGCGAGGAGTCCGCAAGGCTGTC 1656
Db |||||
QY 4320 TACGGTACTACGGCTATGCAAGGCGCGCTGTACTGCAAGGAGTGTGAGAGCGGTA 4379
Db |||||
QY 1657 ACGGCTGGGTAGGCACCTCATACGACCGCATCAAC---ATAGCTCGTAACTAGGC 1713
Db |||||
QY 4380 ACGGCTGGGAGGAGGTACATAACGATGACCATCAAGGAGTAGAGAAAGTACGGC 4439
QY 1714 CTCAGGTGATCTACGGTGCACAGATTGCTCTTGTGACCTATGATCCGGAGAGGTG 1773
Db |||||
QY 4440 TTTAAGGTATCTACAGCGACACCGAGGATTTTTCACCAATACCTGGAGCGGATGCT 4499
QY 1774 GAAATTTTCATCAAAATTTATAAGGAGGAGCTG----- 1806
Db |||||
QY 4500 GAAACGCTCAAAAGAGGCTATGAGTTCCTCAACTATATCAACGCCAACTCCGGGC 4559
QY 1807 GGGTTCGAATCAAGCTAGAGAGGTGTACAAACGCTTATTTTACAGAGGCTAAGAG 1866
Db |||||
QY 4560 GCGCTTGAGCTCGAGTACGAGGCTTCTACAAACGCGCTTCT---CGTCAGGAGAG 4616
QY 1867 AGGTACGCTGGCCTTCTCGAGGACGAGCTATAGATTGTCGGTTCGAGGCTGTACGT 1926
Db |||||
QY 4617 AAGTATGCGGTGATACGAGGAGGAGGATACCAACCGCGGACTTGGAGTTGTGAGG 4676
QY 1927 GCGATTTGTTGAACTCGCCAGGAGGTTACAGTAAAGTGTTCGAAATAGTATTGAAG 1986
Db |||||
QY 4737 GAGCGTTCAGGAGGAGGCTGAGGATAGTCAAGAGGTTTACCGAAAGCTGAGCAAG 4796
QY 2047 GGCAGGTTCCCATAGAGAGCTTGTAACTCGAGAGCCCTTAGTAAGGCTTTGAGGAG 2106
Db |||||
QY 4797 TACGAGTTCGCGGAGAGAGCTGGTGTATCCACGAGCAGATAACGAGGAGTTTAAAGGAC 4856
QY 2107 TACCAACGAGGAGGACACACACGCTGTTGACGAGGAGGATGCTGTACGAGGCTACCGG 2166
Db |||||
QY 4857 TACAGGCAACCGGTCCTCCACGTTGCGGTTCCCAAGAGGTTGCCCGGAGGAGTCAAA 4916
QY 2167 GTAGCCCAAGGCAAGATAGGATATGTAATAGTAGAGGCTGGTGGCCGCTATCAGTCAA 2226
Db |||||
QY 4917 ATACGCGCTGGAACCGGTGATAAGCTCATCGTCTCAAGGCTCTCGGAGGATAGCGGAC 4976
QY 2227 AGAGCATGCGCATACCTTCATGCTCAAGGATCCTAGCCAGATAGAGCTGACC---TACTAT 2283
Db |||||
QY 4977 AGGCGATACCGTTCCAGCGAGTTCCGCCGAGCAAGACAGTACGATGCGGAGTACTAC 5036
QY 2284 GTTGACCAACCAATCATCCCGGCTGCATTGAGAAATACCTGGGCTACTTTGGCATCCGAG 2343
Db |||||
QY 5037 ATTGAGAACCGGTTCTCCCGCGCTTGAGAGATCTTGAGAGCCTTCGTTACCGCAG 5096
QY 2344 AAGAGCTG 2352
Db |||||
QY 5097 GAAGACCTG 5105
Db |||||

RESULT 14

US-09-363-095-6
; Sequence 6, Application US/09363095
; Patent No. 6187573
; GENERAL INFORMATION:
; APPLICANT: IMANAKA, Tadayuki

; APPLICANT: TAKAGI, Masahiro
; APPLICANT: MORIKAWA, Masaaki
; TITLE OF INVENTION: DNA Encoding A Thermostable DNA Polymerase
; FILE REFERENCE: 2418/11
; CURRENT APPLICATION NUMBER: US/09/363,095
; EARLIER FILING DATE: 1999-07-30
; EARLIER APPLICATION NUMBER: US 09/073,259
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: US 08/656,005
; EARLIER FILING DATE: 1996-05-24
; EARLIER APPLICATION NUMBER: JP 134096/95
; EARLIER FILING DATE: 1995-05-31
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: WordPerfect 6.1 Windows
; SEQ ID NO 6
; LENGTH: 5339
; TYPE: DNA
; ORGANISM: Hyperthermophilic archaeon
; FEATURE:
; OTHER INFORMATION:
US-09-363-095-6

Query Match 5.4%; Score 131.4; DB 3; Length 5339;
Best Local Similarity 52.2%; Pred. No. 6.7e-31;
Matches 412; Conservative 0; Mismatches 341; Indels 36; Gaps 4;
QY 1597 TACGGCTACATGCTGGAGCGCGCTAGGTGGTATTGCGAGGAGTCCGCAAGGCTGTC 1656
Db |||||
QY 4320 TACGGTACTACGGCTATGCAAGGCGCGCTGTACTGCAAGGAGTGTGAGAGCGGTA 4379
Db |||||
QY 1657 ACGGCTGGGTAGGCACCTCATACGACCGCATCAAC---ATAGCTCGTAACTAGGC 1713
Db |||||
QY 4380 ACGGCTGGGAGGAGGTACATAACGATGACCATCAAGGAGTAGAGAAAGTACGGC 4439
QY 1714 CTCAGGTGATCTACGGTGCACAGATTGCTCTTGTGACCTATGATCCGGAGAGGTG 1773
Db |||||
QY 4440 TTTAAGGTATCTACAGCGACACCGAGGATTTTTCGCAATACCTGGAGCGGATGCT 4499
QY 1774 GAAATTTTCATCAAAATTTATAAGGAGGAGCTG----- 1806
Db |||||
QY 4500 GAAACGCTCAAAAGAGGCTATGAGTTCCTCAACTATATCAACGCCAACTTCGCGGC 4559
QY 1807 GGGTTCGAATCAAGCTAGAGAGGTGTACAAACGCTTATTTTACAGAGGCTAAGAG 1866
Db |||||
QY 4560 GCGCTTGAGCTCGAGTACGAGGCTTCTACAAACGCGCTTCT---CGTCAGGAGAG 4616
QY 1867 AGGTACGCTGGCCTTCTCGAGGACGAGCTATAGATTGTCGGTTCGAGGCTGTACGT 1926
Db |||||
QY 4617 AAGTATGCGGTGATACGAGGAGGAGGATACCAACCGCGGACTTGGAGTTGTGAGG 4676
QY 1927 GCGATTTGTTGAACTCGCCAGGAGGTTACAGTAAAGTGTTCGAAATAGTATTGAAG 1986
Db |||||
QY 4677 TACGAGTTCGCGGAGAGGCTGAGGATAGTCAAGAGGTTTACCGAAAGCTGAGCAAG 4736
QY 1987 ACGAGTACGAGTGAACAGGCTGTAGATACGTCAGGAGAGATTGTGAAGAGTTGAGGAG 2046
Db |||||
QY 4737 GACGCTGACGCTCGAGAGGCGCTGAGGATAGTCAAGAGGTTTACCGAAAGCTGAGCAAG 4796
QY 2047 GGCAGGTTCCCATAGAGAGCTTGTAACTCGAGAGCCCTTAGTAAGGCTTTGAGGAG 2106
Db |||||
QY 4797 TACGAGTTCGCGGAGAGAGCTGGTGTATCCACGAGCAGATAACGAGGAGTTTAAAGGAC 4856
QY 2107 TACCAACGAGGAGGACACACACGCTGTTGACGAGGAGGATGCTGTACGAGGCTACCGG 2166
Db |||||
QY 4857 TACAGGCAACCGGTCCTCCACGTTGCGGTTCCCAAGAGGTTGCCCGGAGGAGTCAAA 4916
QY 2167 GTAGCCCAAGGCAAGATAGGATATGTAATAGTAGAGGCTGGTGGCCGCTATCAGTCAA 2226
Db |||||
QY 4917 ATACGCGCTGGAACCGGTGATAAGCTCATCGTCTCAAGGCTCTCGGAGGATAGCGGAC 4976
QY 2227 AGAGCATGCGCATACCTTCATGCTCAAGGATCCTAGCCAGATAGAGCTGACC---TACTAT 2283
Db |||||
QY 4977 AGGCGATACCGTTCCAGCGAGTTCCGCCGAGCAAGACAGTACGATGCGGAGTACTAC 5036
Db |||||

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 14:56:38 ; Search time 21 seconds
(without alignments)
3677.307 Million cell updates/sec

Title: US-10-034-849-2
Perfect score: 4188
Sequence: 1 MTEVFTVLDSSEYVVGKEP.....LKASATQKTLDFLAKSK 803

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76:*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3033.5	72.4	803	2 B56277	DNA-directed DNA p
2	2434	58.1	784	2 E72515	probable DNA-direc
3	2246	53.6	781	2 JC7382	DNA-directed DNA p
4	1581	37.8	781	2 A69312	DNA polymerase B1
5	1287	30.7	764	2 S75407	probable DNA-direc
6	1236.5	29.5	775	2 S35543	DNA-directed DNA p
7	1207.5	28.8	771	2 C75023	DNA polymerase I p
8	1197.5	28.6	775	2 S67920	DNA-directed DNA p
9	985.5	23.1	1312	2 S68593	DNA-directed DNA p
10	955.5	22.8	1235	2 C71210	probable DNA-direc
11	915	21.8	586	2 C69028	DNA-dependent DNA
12	771.5	18.4	1086	2 T40242	DNA polymerase del
13	768.5	18.4	1670	2 S71551	DNA-directed DNA p
14	733.5	18.2	1086	2 T43286	DNA-directed DNA p
15	760.5	18.2	1107	1 A41618	DNA-directed DNA p
16	757.5	18.1	1106	1 A39299	DNA-directed DNA p
17	751	17.9	1038	1 JC5757	DNA-directed DNA p
18	749	17.9	1038	2 T18222	DNA polymerase del
19	743.5	17.8	1105	1 S40243	DNA-directed DNA p
20	741.5	17.7	1088	2 T05731	DNA-directed DNA p
21	731.5	17.5	1084	1 S19661	DNA-directed DNA p
22	718.5	17.2	901	2 R84210	DNA polymerase B1
23	713	17.0	1702	2 S42459	DNA-directed DNA p
24	705	16.8	1094	2 S22573	DNA-directed DNA p
25	691.5	16.5	1097	1 RNBV13	DNA-directed DNA p
26	671	16.0	1634	2 B64410	DNA-directed DNA p
27	649.5	15.5	879	2 A56277	DNA-directed DNA p
28	647.5	15.5	872	2 JC7380	DNA-directed DNA p
29	628.5	15.0	875	2 JC5186	DNA-directed DNA p

30	625.5	14.9	959	2 F72763	probable DNA-direc
31	598.5	14.3	882	2 S23019	DNA-directed DNA p
32	598	14.3	1081	2 T20698	hypothetical prote
33	594.5	14.2	882	2 F90201	DNA polymerase I (
34	579.5	13.8	1462	1 DJH0AC	DNA-directed DNA p
35	578	13.8	1465	2 S45628	DNA-directed DNA p
36	569	13.6	787	2 E82227	DNA polymerase II
37	564.5	13.5	844	2 T31321	DNA-directed DNA p
38	527.5	12.6	787	2 G83410	DNA polymerase II
39	526	12.6	3122	2 T17202	DNA-directed DNA p
40	524.5	12.5	1015	1 DJBE2L	DNA-directed DNA p
41	522.5	12.5	1339	1 S20052	DNA-directed DNA p
42	517	12.3	1505	2 S28079	DNA-directed DNA p
43	515	12.3	1009	1 DJBEM2	DNA-directed DNA p
44	511	12.2	1026	2 T03108	DNA-directed DNA p
45	508.5	12.1	1513	2 T28158	probable DNA-direc

ALIGNMENTS

RESULT 1

B56277
DNA-directed DNA polymerase (EC 2.7.7.7) II - Pyrodicticum occultum
C;Species: Pyrodicticum occultum
C;Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 20-Jun-2000
C;Accession: B56277

R;Demori, T.; Ishino, Y.; Doi, H.; Kato, I.

J. Bacteriol. 177, 2164-2177, 1995

A;Title: The hyperthermophilic archaeon Pyrodicticum occultum has two alpha-like DNA pol

A;Reference number: A56277; MUID:95238290; PMID:7721707

A;Accession: B56277

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-803 <DEM>

A;Cross-references: GB:D38574; NID:G807829; PIDN:BA07580.1; PID:G807830

C;Superfamily: herpesvirus DNA-directed DNA polymerase

C;Keywords: nucleotidyltransferase

Query Match 72.4%; Score 3033.5; DB 2; Length 803;
Best Local Similarity 70.9%; Pred. No. 1.6e-176;
Matches 571; Conservative 106; Mismatches 121; Indels 7; Gaps 5;

QY	1	MTEVFTVLDSSEYVVGKEPQVIWGIWIAENGERVWLDSPRPFPYALLAPGADP--KQ	57
DB	1	MTETIEFLLDSSYEILGKEPVVILWGLDQKRVLLDHRFRFPYFIALIARGYEDWVE	60
QY	58	VAQIRALSRKSPITIGVEDDKKVFGRPRRLRTVLPFAVREYRELVKNVGVGVVL	117
DB	61	IAASIRLSVVKSPIIDAKPLDKRYFGRPKAVKITTMIPESVRYREAVKKEGVEDSL	120
QY	118	EDIRPAMRYLIDHDLFPFTWYRVEAPLENKMGFRVDKYLVKSRPEPLYGEALAPT	177
DB	121	EADIRPAMEYLDKRLYPFTVYRIPVEDAGRNPGFRVDKYLKSRPEPLYGEALAPT	178
QY	178	PDLRIILAPDIYVSKGSPRPDPVITAVKTDGDEVLFIAEGKDDKRPREFVYVK	237
DB	179	PMRLVAFDIYVSRGSGSPNPARDPVIIVSLDSEKRLIEAGHDDRRVLEFVEYR	238
QY	238	RYDPDIIYGVNNHFDWPLLRRRTIGIKLDVTRVGAETTSVHGHSVSPGRINVDLY	297
DB	239	AFDPDIYGVNNHFDWPLLRRRTIGIKLDVTRVGAETTSVHGHSVSPGRINVDLY	298
QY	298	DYAEEMPEIKSLBEVAFYLGVMKKSERVIINWHEIFDYWDPPKRPILLOAYRDDVRA	357
DB	299	DYAEEMPEIKSLBEVAFYLGVMKKSERVIINWHEIFDYWDPPKRPILLOAYRDDVRA	358
QY	358	TYGLAEKILPFAIQISYVTGLDQVGAMSGVFRLEWILIRAAFKMKELVPRVRPSET	417
DB	359	TYGLAEKMLPFAIQISYVTGLDQVGAMSGVFRLEWILIRAAFKMKELVPRVRPSET	418
QY	418	YRGAVILEPLRGVHENIAVLDFSSVYVNTIMIKYVGPDTLVRPGEKCGEC-GCWEAPVK	476

```

Db 419 YGAVVLKPLKGVHENVVLDFFSSMYPSIMIKYNVGDPTTVDPSCKPKYGCYVAPVG 478
QY 477 HRRRCPPGFFKTVLRLLELRRAEMKKYPPDSEYLLDEROKALKVLANASYGM 536
Db 479 HRRRCPPGFFKTVLRLLELRRAEMKKYPPDSEYLLDEROKALKVLANASYGM 538
QY 537 GWSGARWYCECAKAVTAGRHILRTAINIARKLGKVIYGDTSLSFVYDPEKVENFIK 596
Db 539 GWSGARWYCECAKAVTAGRHILRTAINIARKLGKVIYGDTSLSFVYDPEKVENFIK 598
QY 597 IIEELGFEIKLEKVKRLFFTEAKRYAGLLEDGRIDIVGEAVRGDMCELAKEVQTKV 656
Db 599 FVEELGFEIKLEKVKRLFFTEAKRYAGLLEDGRIDIVGEAVRGDMCELAKEVQTKV 658
QY 657 VEIVLKTSEVNAKAVVVKVLEKRGKVPKIEKLVTKLSKRLBYTTTEAPVVAARM 716
Db 659 AEIVLNTGVNKAISVIREVVKQREGKVPYKLIIVTKLSKRLBYTTTEAPVVAARM 718
QY 717 LSGYRVSPGDKGYVIVKGGRIQRAMPYFMVKDPSQIDVTYVYDHOIIPAAIRILGY 776
Db 719 KEAGYEVSPGDKGYVIVKGGRIQRAMPYFMVKDPSQIDVTYVYDHOIIPAAIRILGY 777
QY 777 FGITEKKLASATGQKTLDELAKK 801
Db 778 FGVTEKQLKAAATVQSLDFEASK 802

RESULT 2
E72515
probable DNA-directed DNA polymerase APE2098 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hainawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; KDNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A:Reference number: A72450; PMID:99310339; PMID:10382966
A:Accession: E72515
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-784 <KAW>
A:Cross-references: DDBJ:AP000063; NID:G5105654; PID:BA81109.1; PID:G5105797
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE2098
C:Superfamily: herpesvirus DNA-directed DNA polymerase

Query Match 58.1%; Score 2434; DB 2; Length 784;
Best Local Similarity 59.1%; Pred. No. 4.2e-140;
Matches 469; Conservative 113; Mismatches 189; Indels 22; Gaps 5;

QY 17 GKPEQVITWGAENGERVVLIDRSPPYFYALLAPGADPKQVQIRALSPKPIICVE 76
Db 3 GSTPVIILWGRGADSRVVFGEPRFYVLPDGSVGLDQLAMIRLSRPSPIISVE 62
QY 77 DDKRYFGRPRVLRTVLPVPAVREYRELKVNVDGVLEADIRFAMRYLIDHDLPFF 136
Db 63 RVRRRFGREVEALKVTLVPAVREYREAVRLGGVLDVLEADIPALRIIDFNLYPM 122
QY 137 TWYRVEAPLENKGVFRVDKVLV-----KSRPEPLVGEALAPTKLPDLRIAFDI 187
Db 123 RWYAEVREAVVPHGYSVDYATLTSGDIREDETRIQEDPLK-----LWMAFDI 172
QY 188 EYVSKQSPRPDPVIVIAVKTDDGDEVLFIAEGDKDRKPIRFEVYVRYDDIIVGY 247
Db 173 EYVSKMPTDKKDPVIMIGLQQAGGTEILEASDRSKVIAQFVERKVSIDPDIVGY 232
QY 248 NNNHFDWPLYLRRARILKLDVTRRVGAEPSTTSVGHVSVFGLNVDLDVYAEEMPEIK 307
Db 233 NQNRFDWPLYLRRARILKLDVTRRVGAEPSTTSVGHVSVFGLNVDLDVYAEEMPEIK 291
QY 308 JKSLVEAVYLGVNKKSERVIVNWEIPDYWDDEPKKPLLLQYARDVRYATYGLAEKILP 367

```

```

Db 292 VKTLEEVADYLVGVKIGERVLTLEWQLGEYWDSSKREILRKYLRLDDVRSTMGAEFLP 351
QY 368 FAIQSLSYVTGLPLDQVGAMSGFRLEWYLIRAAFMKELVNRVERPEETVYRGAIVLEPL 427
Db 352 FGAELSQVSGFLPLDQVMAASVGFLEWFLIREAAKIGELVNRVERSEGRYAGAVLRPK 411
QY 428 RGVHENTAVLDFFSSMYPSIMIKYNVGDPTTVPQSKGCGCWEAPEVYHFRRCPPGFF 487
Db 412 FGVHEDTAVLDFAFASMYPSIMIKYNVGDPTTVPQSKGCGCWEAPEVYHFRRCPPGFF 471
QY 488 KTVLERLELKRVRKRAEMKKYPPDSEYLLDEROKALKVLANASYGMGPHARWYCRE 547
Db 472 KTLERLELKRVRKRAEMKKYPPDSEYLLDEROKALKVLANASYGMGPHARWYCRE 531
QY 548 CAKAVTAGRHILRTAINIARKLGKVIYGDTSLSFVYDPEKVENFIKIEELGFEIK 607
Db 532 CAEAVTAGRHILRTAINIARKLGKVIYGDTSLSFVYDPEKVENFIKIEELGFEIK 591
QY 608 LEKVYKELFFTEAKRYAGLLEDGRIDIVGEAVRGDMCELAKEVQTKVVEIVLKTSEVN 667
Db 592 VDKYRVRFFTEAKRYAGLLEDGRIDIVGEAVRGDMCELAKEVQTKVVEIVLKTSEVN 651
QY 668 KAVEYVRKIVKLEEGKVPKIEKLVTKLSKRLBYTTTEAPVVAARMKLSAGYRVSPGD 727
Db 652 EAVDYVRNIEKLRGQVDMKLVTKLTPRPSMYEARQPHVTAALLMERAGIKVEPGA 711
QY 728 KIGYVIVKGGRIQRAMPYFMVKDPSQIDVTYVYDHOIIPAAIRILGYGITEKKLAS 787
Db 712 KIGYVIVKGGRIQRAMPYFMVKDPSQIDVTYVYDHOIIPAAIRILGYGITEKKLAS 770
QY 788 ATGQKTLDFELAK 800
Db 771 GR-QSTLDFEVR 782

RESULT 3
JC7382
DNA-directed DNA polymerase (EC 2.7.7.7) B3 - Sulfurisphaera ohwakuensis
X:Alternate names: DNA polymerase alpha, beta, gamma; DNA polymerase 1, II, III
C:Species: Sulfurisphaera ohwakuensis
C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 31-Dec-2000
C:Accession: JC7382
R:Iwai, T.; Kuroiwa, N.; Itoh, Y.H.; Kimura, N.; Horiuchi, T.
DNA Res. 7, 243-251, 2000
A:Title: Sequence analysis of three family B DNA polymerases from the thermoacidophilic
A:Reference number: JC7382
A:Accession: JC7382
A:Molecule type: DNA
A:Residues: 1-781 <IWA>
A:Cross-references: DDBJ:AB032376
C:Comment: This enzyme has both 3'-5' exonuclease and polymerase activities, and plays
C:Genetics:
A:Gene: B3
C:Keywords: DNA replication; exonuclease; metal binding; nucleotidyltransferase

Query Match 53.6%; Score 2246; DB 2; Length 781;
Best Local Similarity 54.3%; Pred. No. 1.1e-128;
Matches 433; Conservative 141; Mismatches 203; Indels 20; Gaps 10;

QY 6 FTVLDSYEVGKPEQVITWGAENGERVVLIDRSPPYFYALLAPGADPKQVQIRAL 65
Db 5 FFLDPSYDVNENKPVYIIVWIDKGNRVLLKPRFYFALVDSDSINIDIRKEIKL 64
QY 66 SRPKSPIIGVDDKRYFGRPRVLRTVLPVPAVREYRELKVNVDGVLEADIRFAM 125
Db 65 SRPKSPIIGVDDKRYFGRPRVLRTVLPVPAVREYRELKVNVDGVLEADIRFAM 124
QY 126 RYLDHDLDPFFTWYRVEAPLENKGVFRVDKVLVKSRRPELYGSAALATKLPDLRIAF 185
Db 125 RYLDHDLDPFFTWYRVEAPLENKGVFRVDKVLVKSRRPELYGSAALATKLPDLRIAF 177
QY 186 DIEVYSKQSPRPDPVIVIAVKTDDGDEVLFIAEGDKDRKPIRFEVYVRYDDIIV 245

```


QY	61	RIRALS	RKSPKSPILIGVEDDKRKFKYRPRVRIRI	RTVLDEAVREYRELXVNDGV	VEDLEAD	120
DB	57	AATSTR	REYIKLKSPEKAQLKTLGREVEGYIVY	AAHPHQHPKLDKLSQFG--	DYREAD	113
QY	121	IRFAMRYLIDH	LPFTWYRVEASPLENMQG----	FRVDKVLVKSRRPELYGEALAPT	175	
DB	114	IPFANYLIDKDL	ACWQJAGEB-----	XQGGVIRS	KIEKV-----	EXIPRM 157
QY	176	KLPDRIILAF	LEVIVSKGSRPRSPDRPVIVIAKTD	DGDEVLF	FAEGKDDRKIRFPEVY	235
DB	158	EFPELKMVL	FOCEMULSSFGMEPEKOPITIVSVK	TNDDDDIILTGG--	DERKTIISDFVKL	214
QY	236	VKRYDPDIIV	GVNNHNPWPVLLRARITIGLKDV	RRVGAEPITTVSH--	GHSVSVPEGLN	293
DB	215	IKSYDPDIIV	GVNQDAFOWPVLKREARNIPLD----	VGRDGSNV	VRGGRKPIVGRLN	270
QY	294	VDLYVAEMPEIK	TSLEBVAEYGVNKKSKSRVINW	WEIPDYWD	DPKRPILLLOYARD	353
DB	271	VDLYDIAMRIS	DIKITKULENVAEPLGT--	KIEIADIEAKDIYRWSRGEKEK-	VLNVABQ	327
QY	354	DVRATVGLAE	KILPFAIOLSVVTGLPDVGAMSG	FELEWYLIRAAFKMBSLVPNRVER	413	
DB	328	DAINYILAKELL	PHYELSKWIRLPVDDVTRMG	RQGVDMWLLSEAKKICEIAPNPPH	387	
QY	414	PBETVRGAI	VLBPGLGVHENTAVIDFSSWY	PNIMIKYNVGPDTLVRPEK	KGBCGGCWAP	473
DB	388	-AESYEGAF	VLPERGLEHNVACLFASMY	PSIMAFNISPDTY-----	GCRD-DCYEAP	440
QY	474	EYVCHFRPC	POFFKTVLLELLEKRVRAEAM	KVYPPDSPEYRILLDEROKALKVLN	ANASY	533
DB	441	EVGHFRKSP	POFFKRIELMLIEKREELKVEL	KMLSPESSEYKLLDIKQOTL	KVLNTNSFY	500
QY	534	GYNGWSGAR	WYCRECAKAVTAWGRHLIR	TAINIARKLGLKVIVGDT	DSLFTV--	YDPEX 590
DB	501	GYMGWNLAR	WYCHPCAETATWAGSHFIR	TSAKIAESMGFKVLXGDT	DSIEVTKAGMTKED	560
QY	591	VENFIKILKE	LFGFIKLEKVKYKRLFFTEAK	RKYAGLLEDGRIDIVGFEAVR	GDWCELAK	650
DB	561	VORLUDKH	SELUPTQIODEVYESIAIFPVE-	KRRYAGLITDGLRVVXGLE	VRRGDWCELAK	619
QY	651	EVQTKVVEI	VLKTSRNVKAVYVRKIVKEL	BEGKVPIEKLVINWKTLS	KRLSEYTTAPHV	710
DB	620	KYQREVIEW	ILKEKNEKALSLVDRVILR	IKEGKVSLEEVVIYKGLTK	PKSPKYESMOAHV	679
QY	711	VRAKMLSG	VRVSPGDGKIVYVKGGRIS	ORAWPYMUKD-----	PSQ	755
DB	680	KALKAREM	GIIVPVSSKIGIVIVKSGNI	UGDRAIPIDLIEDP	DGEMLRITKWSG	IEIKK 739
QY	756	INDVYTVDR	QIILPAALRILGVFGITEK	KKLARSATQKTLF	DFLA	799
DB	740	LQKDYIND	NOIIPSVLIRLEFVGYTEAS	LKGSS--OMSLDS	FFFS	761

RESULT 5

A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiaeh, P.; Kaine, B.P.; Sykes, S. Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Ignicoccus hospitalis*.
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: A69312
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-781 <KLE>
A:Cross-references: GB:AE001070; GB:AE000782; MD:G2689393; PIDN:AAB90741.1; PID:G265013
C:Superfamily: herpesvirus DNA-directed DNA polymerase

Query Match 37.8%; Score 1581; DB 2; Length 781;
Best Local Similarity 42.5%; Fred.No. 2.9e-98;
Matches 350; Conservative 133; Mismatches 273; Indels 68; Gaps 18;
Qy 1 MTEVFTVLDSSVEYVGKPEQVIWIGIAENGERVWLIDRSFYPYVALLACPAQPKVQAQ 60
Db 1 MERVSEGLIDADYETIGGKAVVYRWCKDDG--TFVADVNYDFYFVFI--GVSDRIILN 56

	Query Match	Best Local Similarity	DB 1	DB 2	Length	781;
Qy	1	42.58;	Prod No 2.9e-98			
Db	1	Matches 350;	Conservative 133;	Mismatches 273;	Indels 68;	Gaps 18;

	Query Match	Best Local Similarity	DB 1	DB 2	Length	781;
Qy	1	42.58;	Prod No 2.9e-98			
Db	1	Matches 350;	Conservative 133;	Mismatches 273;	Indels 68;	Gaps 18;

A;Molecule type: DNA
A;Residues: 1-775 <DEM1>
A;Cross-references: EMBL:DL2983; NID:g216917; PID:BAA02362.1; PID:g216918
A;Accession: S44596
A;Molecule type: protein
A;Residues: 1-12 <DEM2>
C;Genetics:
A;Gene: pol
C;Superfamily: herpesvirus DNA-directed DNA polymerase
C;Keywords: DNA binding; nucleotidyltransferase
F;1-775/Product: DNA-directed DNA polymerase #status experimental <WAT>

RESULT 6
S35543
DNA-directed DNA polymerase (EC 2.7.7.7) - *Pyrococcus furiosus*
C/Species: *Pyrococcus furiosus*
C/Date: 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 20-Jun-2000
C/Accession: S35543; S44596
R/Vemori, T.; Ishino, Y.; Toh, H.; Asada, K.; Kato, I.
Nucleic Acids Res. 21, 259-265, 1993
A/Title: Organization and nucleotide sequence of the DNA polymerase gene from the archae
A/Reference number: S35543; MUID:93181200; PMID:8441634
A/Accession: S35543

RESULT 7

C75023
 DNA polymerase I PAB1128 - Pyrococcus abyssi (strain Orsay)
 C:Species: Pyrococcus abyssi
 C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: C75023
 R:anonymous, Genoscope
 submitted to the EMBL Data Library, July 1999
 A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
 A:Reference number: A75001
 A:Accession: C75023
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-771 <KAW>
 A:Cross-references: GB:AJ248288; GB:AL096836; NID:95458960; PIDN:CAB50625.1; PID:G54591.3
 A:Experimental source: strain Orsay
 C:Genetics:
 A:Gene: polI; PAB1128
 C:Superfamily: herpesvirus DNA-directed DNA polymerase

Query Match 28.8%; Score 1207.5; DB 2; Length 771;
 Best Local Similarity 35.7%; Pred. No. 1.4e-65;
 Matches 292; Conservative 150; Mismatches 285; Indels 91; Gaps 21;

Qy 8 VLDSSYEVGKGPQVITGIAENGERRVVLIDSRPFFYALLAPGADPKQV---AQRIR 63
 Db 2 IIDADYITDCGPIIRIFK-KEGGEKVEYDTRFPIYALLKDDSAIDVKKITAEHG 60

Qy 64 ALSRPKSPIIGVEDDKRYKFORRRVIRITVLPVAVREVELVKNVGVGVLEADIRF 123
 Db 61 KIVR-----ITEVEKVKQKFLGRPIEVWKLYLEHPQDPAIREKIREHPAVVDIPEYDIPF 116

Qy 124 AMRYLIDHDLFPFTWYRVEAEPLNKGFRVYLVKSRPEPLVGLALAPTKLPDLRL 183
 Db 117 AKRYLIDKGL-----TPMEGN-----BELTFL 138

Qy 184 AFDIEVYSKQSPRPDPVIVIAVKTDDGDEVL-----FIAEGKDKRPIRBEVY 235
 Db 139 AVDIETLYHEGE-SFGKGPITIMISYADEGAKVITKSIDLPYVEVVSSEMIKRLVKV 197

Qy 236 VKRYDDPIIVGYNHDFWPYLLRRARILGKLDVTRRVGAEPPTSVHGH---VSVFGR 292
 Db 198 IREKDPDVIITNGDNDFPFLKRAEKLGIKPLGRD-NSEPKQRMGDSLAIEIKGRI 256

Qy 293 NVLDLYDAEEMPEIKIKSLAEVAEYLVGMKSERVINWHEIPIYWDPPKRPPLLOYAR 352
 Db 257 HFDLPVIRRTINLFTYTLAEVYEA- FGSKSEKVIYAH--EIAEAWETGKLEKVAKYSM 313

Qy 353 DDVRATYGLAEKILFPFAIQLSYVTGLPDDQVGMASVGFRLWYLIRAAFKMKELVFNVE 412
 Db 314 EDKAVTELGKEFFPMEAQLARLVGQPPWDVSRSSSTGNLVFWFLRKAYERNELAPNKP 373

Qy 413 RPE-----SYRGAIVLEPLRGVHENVIAVDFSSMYPNIMIKYVNGPDTLVRPGEKGE 466
 Db 374 EREYERKLRESYEGGYKPEKGLMEGIVSLDFRSLPSIIITHNVSPDTLNR--ENCKE 431

Qy 467 CGWEAPVKHFRRCPPGPKFTVLERLLELRKVRAMKKYPPDSPEYRLDRQKALK 526
 Db 432 YDV--APQVGHKFCDFPGFTPSLGNLLEBRQIKKRMKE-SKDPVEKKLLDYRQAIK 488

Qy 527 VLNASYGYMGWSGARWYCRECAKAVTAWGRHLIRTAINTARKLGLKVIYDGTDSLFTVY 586
 Db 489 ILANSYGYGYAKARWYCKEASVTANGQYIDLVRERLESFGFKVLIDTDLGYATI 548

Qy 587 D-----PKVENFIKIIKEELG--FEIKLEKVKYKRLFFTEAKRYAGLLEDGRIDV 637
 Db 549 PGAKHEEIKELAFVEYINSKLPGLLELEYEGFYARGFFV-TKKYVALIDECKIVTRG 607

Qy 638 FEAVRGDWCELAKEVQTVKIVLKTSEVKNKAVEYVKIVKELEBEKVPKEKVIWKTLS 697
 Db 608 LEIVRRDMSSEIAKTQAKVLEAILKGNVDKAVKIVKEVTEKSKYIPEPKLVIYEQIT 667

Qy 698 KRLBEYTTAEHVVAAKRMLSAGYRVSPGDKIGYIVKGGGRISQRAWPYFMVK--DPS- 754

Db 668 RPLSEYKATGPHVAVAKRLAAKGVKPGWVIGYIVLRGDPISKRA---IAIESFDPKK 724

Qy 755 -QIDVYVYVDHQIIPALRIILGYFGITEKKLKASATGQ 791
 Db 725 HKYDABYYTINQVLPFAVERILAFAGYRKEDLKYQTKQ 762

RESULT 8
 S67920
 DNA-directed DNA polymerase (EC 2.7.7.7) - Thermococcus sp.
 C:Species: Thermococcus sp.
 C>Date: 17-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 18-Jun-1999
 C:Accession: S67920
 R:Southworth, M.S.; Kong, H.; Kucera, R.B.; Ware, J.; Jannasch, H.W.; Perler, F.B.
 submitted to the EMBL Data Library, January 1996
 A:Description: Cloning, expression and modulation of the 3'-5' exonuclease activity of
 A:Reference number: S67920
 A:Accession: S67920
 A:Molecule type: DNA
 A:Residues: 1-775 <SOU>
 A:Cross-references: GB:U47108; NID:gl197451; PIDN:AAA88769.1; PID:gl197452
 A:Experimental source: strain 90N-7
 C:Superfamily: herpesvirus DNA-directed DNA polymerase
 C:Keywords: DNA binding; nucleotidyltransferase

Query Match 28.6%; Score 1197.5; DB 2; Length 775;
 Best Local Similarity 36.1%; Pred. No. 5.7e-65;
 Matches 301; Conservative 139; Mismatches 296; Indels 97; Gaps 24;

Qy 8 VLDSSYEVGKGPQVITGIAENGERRVVLIDSRPFFYALLAPG---DPRQV-AQRIR 63
 Db 2 ILDTDITENGKRVIVFK-KENGFKIYDIFRFYFALLKDDSAIDVKKVIAKRG 60

Qy 64 ALSRPKSPIIGVEDDKRYKFORRRVIRITVLPVAVREVELVKNVGVGVLEADIRF 123
 Db 61 TVVKVXR---AEKQKFLGRPIEVWKLYLEHPQDPAIRIRAHPAVDIYEDIPF 116

Qy 124 AMRYLIDHDLFPFTWYRVEAEPLNKGFRVYLVKSRPEPLVGLALAPTKLPDLRL 183
 Db 117 AKRYLIDKGL1-----PMEGD-----BELTML 138

Qy 184 AFDIEVYSKQSPRPDPVIVIAVKTDDGDEVLFTAECKD-----RKPIREVF 233
 Db 139 AFDIETLYHEGE-EFTGTILMISYA--DGSARVITWKKIDLPYVDVSTSEMIKRL 195

Qy 234 EYKRYDDPIIVGYNHDFWPYLLRRARILGKLDVTRRVGAEPPTSVHGH---VSVFG 290
 Db 196 RVVREKDPDLITNGDNDFAYLKRCBELGKIFTLGRD-GSEPKIQRMGDRFAVEVG 254

Qy 291 RUNVDLYDAEEMPEIKIKSLAEVAEYLVGMKSERVINWHEIPIYWDPPKRPPLLOY 350
 Db 255 RIHFDPVIRRTINLFTYTLAEVYEA-VFGPKSEKVIYAB--EIAQAMESGGLSERVARY 311

Qy 351 ARDDVRATYGLAEKILFPFAIQLSYVTGLPDDQVGMASVGFRLWYLIRAAFKMKELVFN 410
 Db 312 SMEDAKVTYELGREFPMEAQLSRILGQSLWDVSRSSSTGNLVFWFLRKAYKRNELAPNK 371

Qy 411 VERPEPT-----YRGAIVLEPLRGVHENVIAVDFSSMYPNIMIKYVNGPDTLVRPGEK 465
 Db 372 PDERELARRRGVYAGYVKEPERGLWDNIVYDLFRSLPSIIITHNVSPDTLNR----- 426

Qy 466 ECGCHE--APSVKHFRCPPGPKFTVLERLLELRKVRAMKKYPPDSPYRLDERQ 522
 Db 427 --GCKEYDVAPEVGHKFCDFPGFTPSLGLDLEBRQIKRKOKA-TVDPLEKKLLDYRQ 483

Qy 523 KALKVLNASYGYMGWSGARWYCRECAKAVTAWGRHLIRTAINTARKLGLKVIYDGTDS 581
 Db 484 RAIKILANSFYGYAKARWYCKEASVTANGREYIEMVIREEEKFGFKVYADTGG 543

Qy 582 LFVTV---DPE-----KVENFIKIIKEELG--FEIKLEKVKYKRLFFTEAKRYAGLEDR 632
 Db 544 LHATIPGADAETVKKKAKEFLKVINPKLPGLLELEYEGFYRGFFV-TKKYAVIDEBK 602

QY 633 IDVGFANVAGNCELAKEVQTKVIVLKTSEVNAKAVYVVKVLEEGKVPTEKVI 692
 Db 603 ITRGLRIVRDMSEIAKETQARVLEAILKHGDVEAVRIVKEVTEKLSKYVPEKVI 662
 QY 693 WKTLSKRELEVTAEAPHVAAKRLSAGVRVSGDKIGVIVKGGSRISORAWPFMVKD 752
 Db 663 HEQITRDRDYKATGPHVAVAKELAARGVKIREGTIVISIVLKSGRIGDRAIPDEF-D 721
 QY 753 PS--QIDVYVYVHQIIPALRILGYFGITEKKLKASATGQKTLFDFLAKSK 803
 Db 722 PTKHRYDAEYIENQVLPVERILKAFGYRKEDRYQTKQVGLGAWLAKVKGK 774

RESULT 9
 S68593
 N:Directed DNA polymerase (EC 2.7.7.7) Vent, intein containing precursor - Pyrococcus
 N:Contains: DNA endonuclease (EC 3.1.-.-) PI-1; DNA-directed DNA polymerase (EC 2.7.7.7)
 C:Species: Pyrococcus sp.
 C:Date: 24-Aug-1996 #sequence_revision 01-Nov-1996 #text_change 24-Sep-1999
 C:Accession: S68593
 R:Xu, M.Q.; Southworth, M.W.; Merzha, F.B.; Hornstra, L.J.; Perler, F.B.
 submitted to the EMBL Data Library, August 1993
 A:Description: In vitro protein splicing of purified precursor and the identification of
 A:Reference number: S68593
 A:Accession: S68593
 A:Molecule type: DNA
 A:Residues: 1-1312 <XUA>
 A:Cross-references: EMBL:U00707; NID:G436492; PIDN:AAA67130.1; PID:G025735
 R:Xu, M.Q.; Southworth, M.W.; Merzha, F.B.; Hornstra, L.J.; Perler, F.B.
 Cell 75, 1371-1377, 1993
 A:Title: In vitro protein splicing of purified precursor and the identification of a bra
 A:Reference number: S68581; MUID:94094330; PMID:8269515
 A:Contents: annotation
 R:Xu, M.Q.; Comb, D.G.; Paulus, H.; Noren, C.J.; Shao, Y.; Perler, F.B.
 EMBO J. 13, 5517-5522, 1994
 A:Article: Protein splicing: an analysis of the branched intermediate and its resolution b
 A:Reference number: S52065; MUID:95080235; PMID:7988548
 A:Contents: annotation; self-splicing mechanism
 C:Function: <NUCL>
 A:Description: nucleotidyltransferase
 A:Note: DNA-directed DNA polymerase Vent
 C:Function: <ENDO>
 A:Description: endonuclease; hydrolase
 A:Note: DNA endonuclease PI-PspI
 C:Superfamily: hypothetical protein PH0202
 C:Keywords: DNA replication; endonuclease; hydrolase; nucleotidyltransferase; protein sp
 F:1-491/1030/1312/Product: DNA-directed DNA polymerase Vent #status predicted <MAT1>
 F:1-491/Domain: DNA-directed DNA polymerase Vent extein 1 #status predicted <XT1>
 F:492-1029/Product: DNA endonuclease PI-1 (pol Vent intein 1) #status predicted <MAT2>
 F:1030-1312/Domain: DNA-directed DNA polymerase Vent extein 2 #status predicted <XT2>
 F:491-1030/Cross-link: peptide (Ala-Ser) #status experimental

Query Match 23.1%; Score 965.5; DB 2; Length 1312;
 Best Local Similarity 21.9%; Pred. No. 1.5e-50;
 Matches 300; Conservative 154; Mismatches 282; Indels 635; Gaps 23;

QY 8 VLDSSEYVKEPQVITWINGAENGRVLLIDRSFRYFALLAPGADPKQV-----AQRTR 63
 Db 2 ILDDYITEDGKPIIRIFK-KENGEPKVEYDRNFRFYIALLKDDSQIDVEVRKITAHRG 60
 QY 64 ALSRPKSPITGVEDKRYKQRRRLRIRTVLPEAVREYRELKVNVDGVLEADIRF 123
 Db 61 KIVR-----LDAEKVRKKPLGRPIEVWRLYFEPQDVPAIRDKIRHSADVILFEYDIPF 116
 QY 124 AMRYLIDHDLFPFTWYRVEAPLENKMGFRVDVLYLKSRPEPELYGEALLPTKLPDLRL 183
 Db 117 AKRYLIDKGLI-----PMGED-----BELKXL 138
 QY 184 AFDIEVYSKQSPRPDPVIVIAVKTDDGDEVL-----FIAEGKDRKPIREFVEY 235
 Db 139 AFDIETLYHEGE-EPKAGPIIMISYADEBAKVITWKKIDLPYFVYVSSEREMIKRFLKV 197

QY 236 VKRYDDIIVYNNHFDMPYLLRRARILGKLVTRRVGAEPPTSVHGH---VSVQRL 292
 Db 198 IREKDPVITNGDSFDLPYLVGAELGKILPLGRD-GSEPKMQLRGLDMTAVEIKRI 256
 QY 293 NVLDYVABMBPBIKLSLEVAEYLVGVNKKSERVLIINWWEIFDYDDPKRPLLLQYAR 352
 Db 257 HFDLYHVRITNLPTTLEAVYAI-FGKPKKYAH--ELABAWETGKLERVAKYSM 313
 QY 353 DDVATYGLAEKTLPPAIOLSVTGLPLDQVGAMVGPRLKWLIRAAFKMKELVPRKVE 412
 Db 314 EDKAVTYELGREFFMELQSLVGLQPLMDVSRSTGNLWYLLRKYARNEELAPKPD 373
 QY 413 RPS-----ETVRGAIVLEPLRGVHENIAVLDFSSWPNIMIKYVNGDITLVPRGCGE 466
 Db 374 EREYERLRSTAGGVYKPEKGLWGLVSLDFRSYFSIIITHVNSDITLNR- 427
 QY 467 CGCWE---APEVKHFRRCPPGFFKTVLERLELRKVRVRAEMKKYPPDPSPYRLIDRQK 523
 Db 428 -GCREYDVAPEVGHKFCDFGPGFISLLKGLDDEQEIKRNVKA-SKDPIEKMLDYQR 485
 QY 524 ALKVL- 529
 Db 486 AIKILANSILPEWVPLIKNGKVKIFRIGDFVGLMKANQGKVKTKGTDEVLVAGIHAF 545
 QY 530 529
 Db 546 SPDRKSKARVMKAVIRHRYSGNVYRIVLNSGRKITITGHSILFVYVNGDLVEATGED 605
 QY 530 529
 Db 606 VKIGULLAVPRSVNLPKPERLNIVELLNLSPBETDIIITPVKGRKNFPKGMRLTR 665
 QY 530 529
 Db 666 WIFGEKRYTASRYLRHLENLYELRKIGYDIIDKSGLEKYRTLYEKLVDDVRYNGK 725
 QY 530 529
 Db 726 REYLVFNAVDRVISLMPBEELKSWRIGTRNGFRMGTFVIDEDFAKLGYVYSEGSARK 785
 QY 530 529
 Db 786 WKQVTGWSVTVLYNENDEVLDDMEHLAKKFFGKVRGKNVVEIPKQWYIIPESLCGT 845
 QY 530 529
 Db 846 LAENKRVPEVIFTSSKGVWAFLEGYFVGIGDVHPKRVLRSTKSELLVNGVLLNSLG 905
 QY 530 529
 Db 906 VSAIKLYGDSGVTVTVNEELKFTVYRKKNVTHSHIVPKILKETFGKVFQKNISYKFF 965
 QY 530 529
 Db 966 RELVENGKLDREKAKRIEMLLNGDIVLDRVBEIKREYDGYVYDLSVDEENFLAGFGFL 1025
 QY 530 ---NAGYGVNCSGABWYCEBACAKVATNGRHLIR-TAINTARKLGLKVIYDITDLSFT 585
 Db 1026 YAHNYSYGVYGAARWYCKEACSVTANGREYIIFVRKELEKFGFKVLYITDITDGLYAT 1085
 QY 586 Y---DPEKVE---NFIKIIEELG--FRIKLEKVVKRLFFTEAKKRVAGLEGDIDIV 636
 Db 1086 IPGAKPEEIKKALEFVDYINAKLPGLLEVEYGVYVGFV-TKKYALDDESGKITR 1144
 QY 637 GFNAVGDWCELAKEVQTKVIVLKTSEVNAKAVYVVKVLEEGKVPTEKVI 696
 Db 1145 GLSIVRDMSEIAKETQAKVLEAILKHGDVEAVRIVKEVTEKLSKYVPEKVI 1204
 QY 697 SKLEVTTEAPHVAAKRLSAGVRVSGDKIGVIVKGGSRISORAWPFMVKD---- 752
 Db 1205 TRPEHETKAIGPVAVAKRLAARGVKVRPGWIGVIVLKGDFGFSKRA---ILAEFDLR 1261
 QY 753 PSQIDVYVYVHQIIPALRILGYFGITEKKLKASATGQKTLFDFLAKSK 803

DNA-directed DNA polymerase (EC 2.7.7.7) delta chain - fission yeast (Schizosaccharomyces N; Alternate names: DNA polymerase delta
C; Species: Schizosaccharomyces pombe
C; Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C; Accession: T43266
R; Pignede, G.; Bouvier, D.; de Recondo, A.M.; Baldacci, G.
J. Mol. Biol. 222, 209-218, 1991
A; Title: Characterization of the POL3 gene product from Schizosaccharomyces pombe indica
A; Reference number: S19661; MUID:92071954; PMID:1960723
A; Accession: T43266
A; Status: preliminary; translated from GB/EMBL/DBJ
A; Molecule type: DNA
A; Residues: 1-1086 <PIG>
A; Cross-references: EMBL:L07734; NID:gl173383; PIDN:AAA35303.1; PID:gl173384
C; Genetics:
A; Gene: pold
A; Introns: 77/1
C; Superfamily: herpesvirus DNA-directed DNA polymerase
C; Keywords: DNA binding; DNA replication; nucleotidyltransferase

Query Match 18.2%; Score 763.5; DB 2; Length 1086;
Best Local Similarity 27.2%; Pred. No. 2.1e-38;
Matches 245; Conservative 160; Mismatches 320; Indels 177; Gaps 30;

QY 3 EVVFTVLDSSYEVVKGEPQVVIIVGIAENGERVVLDSPRPYFYALLAPGADPKQVQR 62
DB 97 DIIYFOQIDSEFTGSPVSRILEFVTDNGNSILVHVVGFLPYFYVVKAPVGFPEMLERT 156
QY 63 RAL-----SRPKPIIGVEDDKKYGFRGRVRIRITVLPE 98
DB 157 QDLATCNGGVVDHCHIKENLYFGQNEKSPFKI-----FTNPRILSRARNVPER 210
QY 99 AVREYRELKVNVDGVEDLEADIRFAMRYLIDHDLFPFTWVRVEAEPLNKGFRV---- 154
DB 211 GFNFPEELFPVGVGV-TTESNTQVLLRPMIDCVGVNWIHLPAKQYQFRVQNVSCQ 269
QY 155 -----DKVILVSRPEPLGEALPTKLPDLILAFDIEVYSKQG-SRPERDPVIV 205
DB 270 IEAWINYKDLISL-----PAEGQW-----SNAPLRIMSFIDECAGKRGVFPDPSDPIVQ 320
QY 206 IA-VKTDGDEVLFTA-----EGDKRKPIREBEVYVKRVDPIIVGY 247
DB 321 IASIVTYGDSFPFVNVFCVDCSQIVCTQVFEQNAELSSWSKFRVDVDPVLIGY 380
QY 248 NNNHFTWYLLRARIIGI-----KLDVTRVGAETTSVHGH-----VSPVGRLN 293
DB 381 NTCNFDIPYLLDRAKSLRHNPFLGRHNPFPSVAKETTFSSKAYGTRESKTSIPGR 440
QY 294 VLDYAEEMPIKIKSLEEV-AEVLGVWKKSERVIINWHEIPDYD---DPKRPILLQ 349
DB 441 LMLQVMQRDFKLSYSLNAVCSQFLGEQED-----VHYSITDLQNGTADSRRL--LAI 494
QY 350 YARDVVRATYGLAEKILPFA--IQLSYVTGLPLDOVGAMSVGRLEWYLIRAAFPKELV 407
DB 495 YCLKDAYLPQRLMDKLMCFVNVTENARVTGVPFNFLARGQIKVISQLFCKALQHDIV 554
QY 408 PN-RVERBEETRGAIIVLEPIRGVHEN-IAVLDSSMYPNIM----- 447
DB 555 PNIRVNGTDEQEGATVIEPIKGYVDTEIATLDFSSLYPSIMQAHNLCTVTLSDNTAEL 614
QY 448 -----IKYNGP--DTLVRGCKGCGCEGCEAPVKRFRCPFGPKFTVLEKLELRK 499
DB 615 LKLQKQVDYVTFNGDYVFKPHV-----KGLPIITLADLLNARK 654
QY 500 RYRAEMKVPDPSPEYRLIDRQALKVLANASYGTMGSGARWYCRACAKATANGHL 559
DB 655 KAKADLKK-ETDPFKAVLDGQLALKVSAKSVYGTGATNGRPLFCLAISSSVTSYGRQM 713
QY 560 INTANIRAKL-----GLKVIYGTDSLFVTPDEKVENFKIKIEELGF----- 604
DB 714 IETKQVVEKRYENGYSHDAVVIYGTDSVMVKGKVTLPPEAMKLGEEANVSDQFP 773
QY 605 ----EIKLEKVKRLFTFAKKRYAGLL-----EDGRIDIVGEAVRGDWCELAKEVQTKV 656

RESULT 15

A41618

DNA-directed DNA polymerase (EC 2.7.7.7) delta catalytic chain - human

C; Species: Homo sapiens (man)

C; Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C; Accession: A41618; S35455

R; Chung, D.W.; Zhang, J.; Tan, C.K.; Davie, E.W.; So, A.G.; Downey, K.M.

Proc. Natl. Acad. Sci. U.S.A. 88, 11197-11201, 1991

A; Title: Primary structure of the catalytic subunit of human DNA polymerase delta and

A; Reference number: A41618; MUID:92107916; PMID:1722322

A; Accession: A41618

A; Status: preliminary

A; Molecule type: mRNA

A; Residues: 1-1107 <CHU>

A; Cross-references: GB:M80397; NID:gl81619; PIDN:AAA58439.1; PID:gl81620

R; Yang, C.L.; Chang, L.S.; Zhang, P.; Hao, H.; Zhu, L.; Toomey, N.L.; Lee, M.Y.W.T.

Nucleic Acids Res. 20, 735-745, 1992

A; Title: Molecular cloning of the cDNA for the catalytic subunit of human DNA polymerase

A; Reference number: S35455; MUID:92178967; PMID:1542570

A; Accession: S35455

A; Molecule type: mRNA

A; Residues: 1-29,'H',120-172,'N',174-471,'Y',473-775,'G',777-1107 <YAN>

A; Cross-references: EMBL:M81735; NID:gl81621; PIDN:AAA35768.1; PID:gl81622

C; Genetics:

A; Gene: GDB:FOLD1; FOLD

A; Cross-references: GDB:I129089; OMIM:174761

A; Map position: 19q13.3-19q13.3

C; Superfamily: herpesvirus DNA-directed DNA polymerase

C; Keywords: DNA binding; DNA replication; nucleotidyltransferase; nucleus; zinc finger

Query Match 18.2%; Score 760.5; DB 1; Length 1107;

Best Local Similarity 28.4%; Pred. No. 3.3e-38;

Matches 252; Conservative 145; Mismatches 324; Indels 165; Gaps 31;

QY 17 GKPEQVILMGAENGRVVLIDRSFRPYFVALLAPGADPKQVQRIR-----ALSRRPK--- 69

DB 120 GSVPLTAFQVTDGEGSVCHINGFPAPYFTTAPPGFGPEHMGDLQRENLAISSRSG 179

QY 70 -----SPIGVEDDKK-----YFGR-PRRVLIRITVLPFAVREYRELKVNVDGVEDV--- 116

DB 180 RELTGAVLAVELCSRESFMFGHGHGSPFLRITVALPRIVAPARELLQGIRVAGLSTP 239

QY 117 -----LEADIRFAMRYLIDHDLFPFTWVRVEAEPLNKGFRVDKVL-----VKSRP 164

DB 240 SFAPYANVDFFIRFVVDTDIVGCVNLELPAGKIALRLKEXATQCQLEADVLWSDVWSHP 299

QY 165 EPLYGALAPTCLPDLILAFDIEVYSKQG-SRPERDPVIV-----AVKTDGDEVLFTA 220

DB 300 PEGFWRIAP-----LRVLSFDIECAGRKIGIPEPERDPVIGICSLGRLWGEPEPFLRLA 354

QY 221 -----EGDKRKPIREBEVYVKRVDPIIVGYNHHDWVPLLRRA-- 261

DB 355 LTLRCPAILGAKVQSVKEEDL--LQANSTIRINDPVTIGYNTQNFDLPIYLSRAQT 412

QY 262 -----RIIGIKLDV-----TRRVGABFTTSVHGHSVPGSLNVDLYDAEMPE 305

DB 413 LKVTQTFPFLGRVAGLCSNIRDSFSQSKQGRDRTKV-----VSMVGKRVQMDLQVLLREYK 468

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 14:52:54 ; Search time 17 Seconds

(without alignments)
2221.320 Million cell updates/sec

Title: US-10-034-849-2

Perfect score: 4188

Sequence: 1 MTEVVFVTLDSYEVGKPE.....LKASATQKTLFDPLAKKSK 803

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2434	58.1	784	1	DPO2_AERPE
2	1581	37.8	781	1	DPO1_ARCFU
3	1303.5	31.1	763	1	DPO3_SULSH
4	1287	30.7	764	1	DPO3_SULSO
5	1245.5	29.7	773	1	DPO1_THEGO
6	1236.5	29.5	775	1	DPO1_PYRBU
7	1207.5	28.8	771	1	DPO1_PYRAB
8	1197.5	28.6	775	1	DPO1_THES9
9	986	23.5	824	1	DPO1_METVO
10	965.5	23.1	1312	1	DPO1_PYRSD
11	955.5	22.8	1235	1	DPO1_PYRHO
12	915	21.8	586	1	DPO1_METH
13	845	20.2	1523	1	DPO1_THEFM
14	776	18.5	1671	1	DPO1_PYRKO
15	771.5	18.4	1086	1	DPO1_SCHFO
16	765	18.3	1105	1	DPO1_ORYSA
17	760.5	18.2	1107	1	DPO1_HUMAN
18	757.5	18.1	1106	1	DPO1_BOVIN
19	749.5	17.9	1699	1	DPO1_THEG8
20	749	17.9	1038	1	DPO1_CANAL
21	746.5	17.8	1103	1	DPO1_MESAU
22	743.5	17.8	1105	1	DPO1_MOUSE
23	741.5	17.7	1088	1	DPO1_SOYBN
24	737.5	17.6	1081	1	DPO1_ARATH
25	736.5	17.6	1103	1	DPO1_RAT
26	713	17.0	1702	1	DPO1_THELI
27	705	16.8	1094	1	DPO1_PLARK
28	691.5	16.5	1097	1	DPO1_YEAST
29	672	16.0	1658	1	DPO1_THEHY
30	670	16.0	1634	1	DPO1_METUA
31	658.5	15.7	1829	1	DPO1_THEST
32	655	15.6	1092	1	DPO1_PROME
33	647.5	15.5	872	1	DPO1_SULOH

34 628.5 15.0 875 1 DPO1_SULAC
35 625.5 14.9 959 1 DPO1_AERPE
36 598 14.3 1081 1 DPO1_CAEEL
37 594.5 14.2 882 1 DPO1_SULSO
38 579.5 13.8 1462 1 DPOA_HUMAN
39 578 13.8 1465 1 DPOA_MOUSE
40 568 13.6 1451 1 DPOA_RAT
41 539 12.9 3130 1 DPO2_HUMAN
42 526 12.6 3122 1 DPO2_MOUSE
43 524.5 12.5 1015 1 DPO1_ESV
44 522.5 12.5 1339 1 DPOA_TYRBB
45 515 12.3 1009 1 DPO1_HSVSA

ALIGNMENTS

RESULT 1
DPO2_AERPE
ID DPO2_AERPE STANDARD; PRT; 784 AA.
AC O93746;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase II (EC 2.7.7.7).
GN POLB OR APE2098.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RA Ishino Y., Cann I.K.;
RT "Isolation of the genes encoding two alpha-like DNA polymerases from
RT Aeropyrum pernix."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Onuchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
+ {DNA}(N).
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB017501; BAA75663.1; --
CC EMBL; AP000063; BAA81109.1; --
CC HSP; P56889; ITGO.
CC PIR; E72515; E72515.
CC InterPro; IPR006172; DNA_pol_B.
CC InterPro; IPR006134; DNA_pol_B_dom.
CC InterPro; IPR006133; DNA_pol_B_exo.
CC InterPro; IPR004578; PolZ.
CC Pfam; PF00136; DNA_pol_B; 1.
CC Pfam; PF03104; DNA_pol_B_exo; 1.
CC SMART; SM00486; POLBc; 1.

P95690 sulfolobus
O93745 aeropyrum p
P90829 caenorhabdi
P26811 sulfolobus
P09884 homo sapien
P33609 mus musculu
O89042 rattus norv
O60673 homo sapien
Q61493 mus musculu
P03198 epstein-bar
P27727 trypanosoma
P24907 herpesvirus


```

Db 114 IPFAYRITDKLACHMGIAIEGE-----KQGGVIRSYKIEK-----BRIPRM 157
QY 176 KLPDLRIILAFDJEVYSKQSPERDPVIVIAVKTDDGDEVLFIAEGKDDKRPPIREVEY 235
Db 158 EPELKMVLVDCMLSSFGMPPEKDPPIIVSVKINDDEIILTG-----DERKITSDFVKL 214
QY 236 VKRYDPDIIIVGNHNFDPYLLRRARILIGIKLVDTRRVGAEPTTSVH---GHVSVPGRLN 293
Db 215 IKSVDPIIIVGYNQADFDPYLRKRAERNIPLD---VGRDGSNVVFGGPRPKITGLN 270
QY 294 VDLIYDAEMPIKIKLSLEAVBYLGVWKKSERVINWWEIPDYWDKPKRPLLOIYARD 353
Db 271 VDLIYDIAMRISDIKIKLENVAEFLGT--KIBIADIEAKDIYVNSRGEKEK-VLNYARQ 327
QY 354 DVRAVYGLAEKILPFAIOLSVYTGLPDQGVGMSVGFLEWVLIARAEKMKELPNRVER 413
Db 328 DAINYLIKELPMPHYLSKMRILPVDVTRNGRQKQVMDLLSEAKIGEIAFNPEH 387
QY 414 PEETYGAIVLEPLRGVHENIAVLDFSSMYNIMIKYVNPDTLVRPGEKQCECCWEAP 473
Db 388 -AESYEGAFVLEPRLGHENVACLOFASWYPSIMIAFNISPTTY-----GCRD-DCYAP 440
QY 474 EVKRPFRCPGFFKTVLERLELRKRVRAEMKKYPPDSPEYRLLDERQKALKVLANASY 533
Db 441 EVGHRKFRSPDGFRRILRLMTEKRELKVELKNLSPESSEYKLDIKQKTLKVLTSIFY 500
QY 534 GYMGSGARWYCRECAKAVTAGRHILIRTAIRAKLGVLYGTDSDSLFTV---YDPEK 590
Db 501 GYMGWNLARWYCHPCAETATTAGRHPIRTSAKIAESMGFKVLYGTDIFVTKAGMTKED 560
QY 591 VENFIKIKELGFEIKLEKVKYKLVFFTEAKRYAGLLEDGRIDIVGPEAVRGDWCELAK 650
Db 561 VDRLTKLHEELPIQIEVDYYSIAFFVE-KKRYAGLTEDGRLVYKGLVVRGDCCELAK 619
QY 651 EVQTKVIVILKTSVNKAVERKIVKELEGKVPTEKLVWTKLSKRLBEYTEAPHV 710
Db 620 KVQREIVSVILKEKNEPEKALSIVKQVILRIKEGVSLSEEVYIKGLTKPKSKYESMQAHV 679
QY 711 VAAKMLGAGYRVSQDKIGYVIVKGGGRISORAWPYFMVKD-----PSQ 755
Db 680 KAALKARENGIYVSSKIGYVIVKGGNIGRYPDILDFDGENIRIKTKSGIEIKK 739
QY 756 IDVTYVYDQHIIPAAIRILGYFGITEKKLKASATQKTLFDFLA 799
Db 740 LDKYIDNIQIIPSVLRIERFGYTEASLKGSS--QMSLDSPFS 781

RESULT 3
DPO3 SULSH STANDARD; PRT; 763 AA.
AC Q05706;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase III (EC 2.7.7.7) (DNA polymerase B3).
GN DPO3.
OS Sulfolobus shibatae.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2286;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97252493; PubMed=9098062;
RA Edgell D.R., Klenk H.-P., Doolittle W.F.;
RT "Gene duplications in evolution of archaeal family B DNA
polymerases.";
RL J. Bacteriol. 179:2632-2640(1997).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; U92874; AAB53089.1; --
 DR InterPro; IPR006172; DNA_pol_B.
 DR InterPro; IPR006134; DNA_pol_B_dom.
 DR InterPro; IPR006133; DNA_pol_B_exo.
 DR Pfam; PF00136; DNA_pol_B; 2.
 DR Pfam; PF03104; DNA_pol_B_exo; 1.
 DR SMART; SM00486; POLBc; 1.
 DR PROSITE; PS00116; DNA_POLYMERASE_B; FALSE_NEG.
 KW Transferase; DNA-directed DNA polymerase; DNA replication;
 KW DNA-binding.
 SQ SEQUENCE 763 AA; 88196 MW; 7A7EFD9A5B9305E2 CRC64;

```

Query Match 31.1%; Score 1303.5; DB 1; Length 763;
Best Local Similarity 35.2%; Pred. No. 2.2e-74;
Matches 282; Conservative 164; Mismatches 309; Indels 47; Gaps 14;

QY 6 FTVLSSSYEVGKEPQVIMGIAENGERVVLIDRSFRPYFALLAFGADPKQVAQRAL 65
Db 5 FFILDSFYETKDNIPLIYINSIDDEGNSCVVERNFKPYFVYVVEGND--EIEHNI--- 59
QY 66 SRPKSPILIGVEDDKRKYGPRPRVRIITVLPVAVREYRELKVNQGVDEVLADIRFAM 125
Db 60 -RKNCVELLITKVKYKLVNVVALLVQITPTQIKRCREKIRKINGIKSIFDADIRFTM 118
QY 126 RYLIHDLPFTYRVYAEPLNKMGRVDKVLVKSREPLYGCEALPTKLPDLILAF 185
Db 119 RYSIDFELRFTWFKAEVSEVKUB-GFRACKVILDKILSHYEG-----KIPELRAIGI 171
QY 186 DIEVYKQSGPRERDPVIVIAVKTDDGDEVLFIAEGKDDKRPPIREVEYVKYDPIIV 245
Db 172 DFQIYSKGSNLPRKDPVIVLSLWSKEGSKQPSLDESMDLKIIRFVDVILNDDPIY 231
QY 246 GYNNHFDNPIYLRARILIGIKLQVTRRVGAETTSVGHVSVGRLNVDLYDAYEMPE 305
Db 232 VFDVDFHMKYITRANSLGKIDIGRKIGSEVSQGTGYHYSIGSLNVDLVGLLWNERL 291
QY 306 IKIKSLEEAVALGVNKKSRVILNWEIPDYWDKPKLLOIYADVDVVRATYGLAEKI 365
Db 292 TGHIDILEVANYLGISPK--RDSLNWYEISRYWDDENRDLVKQYSLENAKSIYLLGNFL 349
QY 366 LPFAIQLSVYTGLPDQGVGMSVGFLEWVLIARAEKMKELPNRVERPEETVR-GAIVL 424
Db 350 LSPVSELVKIIGLPDLKLSVASWGNRIEASLIRTAAKSEELIPRMDNPNRSSKIKTVI 409
QY 425 EPLRGVHENIATLDFSSMYPNIMIKYVGPDTLVRPGEKGC-GEWAPEVKKRRPRCP 483
Db 410 EPKIGYSDVYVLDISSVYLSVIRKFNISPDITLVK-----GQDDCYVSTISNYKPKEP 464
QY 484 PGFFKTVLERLLELRKRVRAEMKKYPPDSPEYELDEROKALKVLANASYGVGWSGARW 543
Db 465 SGLYKTFLELSNIQD-----TRKSKVIEELMSSFDYIHWNSRW 505
QY 544 YCRECAKAVTAGRHILIRTAIRAKLGVLYGTDSDSLFTV-YDPEKVENFIKIKKEEL 602
Db 506 YSREIASAVDELSYEIGKLVLDLIKNSGFVILANDFLVFKGSGDKLNEILFKINSIY 565
QY 603 GFEIKLEKVKYKLVFFTEAKRYAGLLEDGRIDIVGPEAVRGDWCELAKEVQTKVVEILK 662
Db 566 DNLKVKRTYRSLLIL-GNDRYAGLLEGDKIDIRIARIGEDRDLCELVRNVKRWVEEILI 624
QY 663 TSVNKAVERVYKVKLEEGKVPTEKLVWTKLSKLEBEYTEAPHVVAARGLMSAGR 722
Db 625 SKDVKAVALVKSANVILKRRGFDIGELITWHIEKDFSEYDKQLPFVVAARAKATQSGYL 684
QY 723 VSPGDKIGYVIVKGGGRISORAWPYFMVKDPSQIDVTYVYDQHIIPAAIRILGYFGITEK 782

```


GN POL OR POLA.
 OS Thermococcus gorgonarius.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Thermococcus.
 RX NCBI_TaxID=71997;
 RN [1]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE=99199230; PubMed=10097083;
 RA Hopfner K.-P., Eichinger A., Engh R.A., Laue F., Ankenbauer W.,
 RA Huber R., Angerer B.;
 RT "Crystal structure of a thermostable type B DNA polymerase from
 Thermococcus gorgonarius";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:3600-3605(1999).
 CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
 CC EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
 CC + {DNA}(N).
 CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
 CC PDB; ITGO; 22-MAR-99.
 DR InterPro; IPR006172; DNA_pol_B.
 DR InterPro; IPR006134; DNA_pol_B_dom.
 DR InterPro; IPR006133; DNA_pol_B_exo.
 DR InterPro; IPR004578; Pol2.
 DR Pfam; PF00136; DNA_pol_B; 1.
 DR Pfam; PF03104; DNA_pol_B_exo; 1.
 DR PRINTS; PR00106; DNAPOLB.
 DR SMART; SM00486; POLBc; 1.
 DR TIGRfams; TIGR00592; pol2; 1.
 DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
 KW Transferase; DNA-directed DNA polymerase; DNA replication;
 KW DNA-binding; Hydrolyase; Nuclease; Exonuclease; Multifunctional enzyme;
 KW 3D-structure.
 FT DISULFID 428 442
 FT DISULFID 506 509
 FT STRAND 2 10
 FT TURN 11 12
 FT STRAND 13 22
 FT TURN 23 24
 FT STRAND 25 31
 FT STRAND 37 42
 FT HELIX 45 47
 FT HELIX 48 51
 FT TURN 52 53
 FT STRAND 55 58
 FT TURN 59 60
 FT STRAND 61 64
 FT STRAND 67 75
 FT TURN 76 77
 FT STRAND 78 86
 FT TURN 90 91
 FT HELIX 92 102
 FT TURN 104 105
 FT STRAND 106 110
 FT HELIX 116 123
 FT TURN 124 125
 FT STRAND 137 144
 FT TURN 152 153
 FT STRAND 157 164
 FT TURN 165 166
 FT STRAND 167 172
 FT TURN 179 180
 FT STRAND 181 183
 FT HELIX 187 201
 FT STRAND 205 208
 FT HELIX 211 213
 FT TURN 214 214
 FT HELIX 215 225
 FT TURN 226 227
 FT TURN 234 235
 FT STRAND 240 243
 FT STRAND 248 251
 FT TURN 253 254
 FT STRAND 256 259

FT HELIX 260 267
 FT HELIX 283 283
 FT TURN 292 301
 FT HELIX 303 304
 FT HELIX 305 337
 FT TURN 338 338
 FT HELIX 341 345
 FT TURN 346 346
 FT HELIX 349 363
 FT TURN 364 365
 FT STRAND 366 366
 FT HELIX 374 379
 FT STRAND 389 390
 FT STRAND 396 405
 FT HELIX 408 415
 FT TURN 416 417
 FT TURN 420 422
 FT STRAND 423 423
 FT TURN 424 424
 FT TURN 426 427
 FT STRAND 431 433
 FT TURN 435 437
 FT STRAND 440 442
 FT HELIX 448 469
 FT HELIX 473 490
 FT TURN 491 492
 FT HELIX 493 498
 FT TURN 500 501
 FT TURN 503 504
 FT HELIX 507 530
 FT TURN 531 532
 FT STRAND 535 540
 FT STRAND 543 547
 FT TURN 549 550
 FT HELIX 553 568
 FT TURN 569 570
 FT TURN 573 574
 FT STRAND 577 590
 FT TURN 591 592
 FT STRAND 593 597
 FT TURN 599 600
 FT STRAND 603 606
 FT STRAND 615 615
 FT HELIX 617 631
 FT TURN 632 633
 FT HELIX 636 651
 FT TURN 652 653
 FT HELIX 657 659
 FT STRAND 662 665
 FT HELIX 670 672
 FT HELIX 679 689
 FT TURN 690 690
 FT TURN 695 696
 FT STRAND 698 704
 FT TURN 710 711
 FT STRAND 714 716
 FT HELIX 717 719
 FT TURN 722 724
 FT STRAND 727 727
 FT HELIX 729 735
 FT TURN 736 737
 FT HELIX 738 740
 FT TURN 741 741
 FT HELIX 742 746
 FT TURN 747 749
 FT HELIX 752 754
 FT TURN 766 767
 FT TURN 771 773
 SQ SEQUENCE 773 AA; 89812 MW; FG7AF04B875FBE44 CRC64;

Query Match 29.7%; Score 1245.5; DB 1; Length 773;
 Best Local Similarity 36.1%; Pred. No. 9.9e-71;
 Matches 300; Conservative 146; Mismatches 289; Indels 93; Gaps 21;


```

Matches 292; Conservative 150; Mismatches 285; Indels 91; Gaps 21;
QY 8 VLDSSYEWGKEPQVITWGAENGVRVLDSPRPYFYALLAPGA---AQRIR 63
D 11ADYITEGKPIIRIFK-KGEFKVEYDRFRPIYVALLKDDSAIDEVKKTAKRHG 60
QY 64 ALSRPSPIIIGVEDDKKYPGRPRRIRTVLPEAVREYRELKVNVDGVLEADIRF 123
D 61 KIVR-----ITEVEKQVKFLGRPIEVWKLVEHPQDPAIREKIREHPAVVIFEDIF 116
QY 124 AMRYLIDHDLFPFTTWVREAEPLNKGFRVQVYLVKSRPELYGEALAPTKLPDLRIL 183
D 117 AKRYLIDKGLI-----TPMEGN-----BELTML 138
QY 184 AFDIEVYSKQSPRPDPVIVIAVKTDDGDEVLFIAEGKDD-----RKPIREFV 233
D 139 AVDIETLYHGE-EFGKGPILMISYA--DGEARVITWKIDLPYVDVSTKEMIKRFL 195
QY 236 VKRYDPIIVGNNHFDWPIYLRARILIKLDVTRRVGAEPPTSVDHGH---VSPGRL 292
D 198 IREKDPDVIITNGDNFDPPYLLKRAEKLGKPLGRD-NSEPKQRMGDSLAVEIKHRI 256
QY 293 NVLDLYAEEMPEIKTSLEVAEYLVGMKKSERVIINWWEIPDYDDPKGRPLLQYAR 352
D 257 HFDLFFVIRRTINLPTTYTLEAVYEAI-FGKSKEKVYAH--EIAEAWETCKGLERVAKYSM 313
QY 353 DDVRATYGLAEKILPFAIQLSYVTGLPDQVGAMSGFRLEWYLIRAAKMKELVPNRVE 412
D 314 EDKAVTEFELCKEFPFMEALQARLVGPVMDVSSSTGNLVWFLRKAYERNELAPNKP 373
QY 413 RPE-----ETRGAIIVLEPLRGVHENIAVLDSSMYPNIMIKYNGPDTLVVRGKCGE 466
D 374 EREYERRLRSEYGGVYKEPEKLGWEGIVSLDPSLYPSIITHNVSPDTLNR--ENCKE 431
QY 467 CGCWAPKVEKHRRCPDPPGFTVLRLELRKVRBAEMKYPSPDSFEVRLDLDERQALK 526
D 432 YDV--APQVGHFRCKDPFGFIPSLNLGNLEERQKIKRMKE-SKOPVEKLLDYRQRAIK 488
QY 527 VLNASYGYMGSGARWYCRCAKAVTANGHILIRTAIARKLGLKVYGTDSLFTY 586
D 489 ILANSYGYGYAKARWYCKEASVTANGQVIDLVRRELSRGFKVLIITDGGIAT 548
QY 587 D-----PEKVENFIKIEELG--FEIKLEKVKRLFFTEAKRYAGLEBDGRIDIVG 637
D 549 PGAKHEBIKALKKVEYINSLKPLGLELEYEGFYARGFV-TKKYALLIDEGKIVRG 607
QY 638 FEAVRGDMCELAKEVQTKVEIVLKTSEYKNAVEYVKRIVKLEGGKVPKLVITWKTLS 697
D 608 LEIVRDMSEIAKETQAKVLEAILKHGNDVAVKIVKSVTEKLSYEIPPEKLVYEQIT 667
QY 698 KRLEEYTTBAPHVAAKRLMSAGYRVSQDKIGYIVKGGGRISORAWPYFWK--DPS- 754
D 668 RPLSEYKALGPHVAVAKRLAANGVKVPGWVIGYIVLRGDPISKRA---IAIEFDPK 724
QY 755 -QIDVTYYVHQIIPALRILGVFGITEKKUKASATGQ 791
D 725 HKYDAEYTYTENQVPAVERILAFAGYRKEDLVQYKTKQ 762

```

RESULT 8

```

DPOL_THES9
ID DPOL_THES9 STANDARD; PRT; 775 AA.
AC Q56366;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN POL OR POLA.
OS Thermococcus sp. (strain 90N-7).
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Thermococcus
OX NCBI_TaxID=103799;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=96224274; PubMed=8643567;
RA Southworth M.W., Kong H., Kucera R.B., Ware J., Jannasch H.W.,
RA Perler F.B.;
RT "Cloning of thermostable DNA polymerases from hyperthermophilic marine
RT Archaea with emphasis on Thermococcus sp. 90N-7 and mutations
RT affecting 3',-5' exonuclease activity."
RL Proc. Natl. Acad. Sci. U.S.A. 93:5281-5285 (1996).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U47108; AA88769.1; -.
DR PIR; S67920; S67920.
DR PDB; 1QHT; 04-JUN-00.
DR InterPro; IPR006172; DNA_pol_B.
DR InterPro; IPR006134; DNA_pol_B_dom.
DR InterPro; IPR006133; DNA_pol_B_exo.
DR InterPro; IPR004578; Pol2.
DR Pfam; PF00136; DNA_pol_B; 1.
DR Pfam; PF03104; DNA_pol_B_exo; 1.
DR PRINTS; PR00106; DNAPOB.
DR SMART; SM00486; POLBc; 1.
DR TIGRFAMs; TIGR00592; pol2; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; 3D-structure.
SQ SEQUENCE 775 AA; 89781 MW; 71525D63C0C9CD9D CRC64;

```

```

Query Match 28.6%; Score 1197.5; DB 1; Length 775;
Best Local Similarity 36.1%; Pred. No. 1e-67;
Matches 301; Conservative 139; Mismatches 296; Indels 97; Gaps 24;
QY 8 VLDSSYEWGKEPQVITWGAENGVRVLDSPRPYFYALLAPGA---AQRIR 63
D 2 ILDTDYITENGKPIRIFK-KGEFKVEYDRFRPIYVALLKDDSAIDEVKKTAKRHG 60
QY 64 ALSRPSPIIIGVEDDKKYPGRPRRIRTVLPEAVREYRELKVNVDGVLEADIRF 123
D 61 TVVKVGR-----AEKQVKKFLGRPIEVWKLVEHPQDPAIREKIREHPAVVIFEDIF 116
QY 124 AMRYLIDHDLFPFTTWVREAEPLNKGFRVQVYLVKSRPELYGEALAPTKLPDLRIL 183
D 117 AKRYLIDKGLI-----TPMEGN-----BELTML 138
QY 184 AFDIEVYSKQSPRPDPVIVIAVKTDDGDEVLFIAEGKDD-----RKPIREFV 233
D 139 AFDIETLYHGE-EFGKGPILMISYA--DGEARVITWKIDLPYVDVSTKEMIKRFL 195
QY 234 EYVKRYDPIIVGNNHFDWPIYLRARILIKLDVTRRVGAEPPTSVDHGH---VSPG 290
D 196 RVREKDPDVIITNGDNFDPPYLLKRAEKLGKPLGRD-NSEPKIQRMGDRFAVVKG 254
QY 291 RLAVDLVDYAEEMPEIKTSLEVAEYLVGMKKSERVIINWWEIPDYDDPKGRPLLQY 350
D 255 RIHFDPVIRRTINLPTTYTLEAVYEAV-FGKPKKRYAE--EIAQWESGEGLERVARY 311
QY 351 ARDDVRATYGLAEKILPFAIQLSYVTGLPDQVGAMSGFRLEWYLIRAAKMKELVPNR 410
D 312 SMEDAKTYELGHEFFMEALQSLRLGSLNDVSRSTGNLVWFLRKAYERNELAPNKR 371
QY 411 VERPEET-----VRGAIIVLEPLRGVHENIAVLDSSMYPNIMIKYNGPDTLVVRGKCG 465
D 372 PDERELARRRGVAGGVYKEPERGLMDNIVLDFRSILYPSIITHNVSPDTLNR----- 426

```

QY	466	EGCWEH---	APVUKHRFRCPGPGFKTVLSELLLELRKRVRAEMKKYPPDPSRYLLDERQ	522
Db	427	--GCKEYDVAPGVGHKFD	PGFTPSLLGDLLEBERQKIRKMKKA-TVDPLEKKLLDYRQ	483
QY	523	KALKVLNANISYMGWSGARWY	CHRECAKAVTAWGRHILRTAI--NIARKUGLKVITYGDDTS	581
Db	484	RAIKILANSFYGYGYARWY	CHRECAESVTAWGREYIENVI	RELBEKFGFKVLVADTDG 543
QY	582	LFVTV---DPE---	KVENETIKITKEELG--FEIKLEKVKRLPPTFEAKRYAGLLEDR	632
Db	544	LHATTPGDAETVKKKAK	EFUKYINPKLPGLLELEYEGFYVRGFFY-TKKKAVIADSEK	602
QY	633	IDIVGFEAVRGDWCALAKEVOT	KVVEIVLKTSEYKNKAVEVVRKI	VKSEBEGKVPTEKVI 692
Db	603	ITTRGLEIVRDNWSIAKETQAR	VLEAILKHGDVEEAVRJVKEVTSEKSYEVPPEKUVI	662
QY	693	WKTLSKRLSEYITEAPHVVAAR	MLSAGYRVSPGDKIGYVIVKGGGRISQRAWPFWMKD	752
Db	663	HEQIETDRDYKATGCPHVA	VAKLAARGVKIRPGTVISIVLXGSGRIGDRAIPADEF-D	721
QY	753	PS--QIDTVYVVDHQIIPAAUR	ILCYGTEKKLKKASATQKTLDFDLAKSK	803
Db	722	PTPKRYDASYIENQVLP	PAVERILKAFGYEKDELRYQTKQVGLGHWLVKVG	774

	Query Match	23.5%; Score 986; DB 1; Length 824;
	Beat Local Similarity 29.0%; Pred. No. 2.1e-54;	
	Matches 255; Conservative 159; Mismatches 288; Indels 178; Gaps 29	
QY	9 LDSSYEVVG-KRPQVILNGIAENGSRVVLIDBSFRPYFALLAPQADPKVQAQIRALS	67
DB	12 IDMYKNGGLKPEINLQKEC-----FKPYFY--VDTSEPKETDYLDGLNQ	57
QY	68 -----PK-----SPIIGVEDDARKYFGPRR	88
DB	58 EIDKLEPEFENNLSLKVDQLITWIEIEKIVSDYILNGKDISEVDFKQK---KERK	114
QY	89 VLRIITVLPVAVREYRELKQVGDVEDLEADIRFAMRYLIDHDLFFFTWYRVAEPLEN	148
DB	115 ICKVYKYPNHVKIIREYKPSFG--KSYEPDIPELRRYMIDQDIVPSAKYS-EDNKIDN	170
QY	149 KMGFRVDKVLVKSREPPLYGALAPTCLDLRLIAFEDIVYSKQGSRPBRDPVIVAV	208
DB	171 -----SIPELNCIAFDMELYCKK-EPNAKQPIIMVNL	202
QY	209 KTDGDEVL-----FIAGKDDRPRIREFYVKRYDPDIIIVGNNHFDWPY	256
DB	203 FSKDYQKVITYKFFENSEYNGCDVYKDEKELIKTIELKQY--DVIYYNGDNDFPY	260
QY	257 LLRRAILGIGLDVTRVQAEFT-----SVGHVSVPGRLVNDVYVAAEEMPEIK	309
DB	261 LKRRANIYIELDFONASQOPIIKISKGINRKSIPGIIHIDLYPIARKLLNLTKY	320
QY	310 SLEBAEYLGVMKSERVILINWETPDYDQPKRPILLQYARDVPAATVGLAEKILPFA	369
DB	321 KLENVQELFKINKE---AVDYGDIPOKWE--TBDTLLRYAYEDALTYTKMGNYFULE	375
QY	370 IQLSYVTGLPQVGAMSGVGFLEWYLIRAAAFKMKELVPNRV-----ERPEETYRGAI	423
DB	376 IMFSRIVQPLDYTRMNSSQWVELLLKRSPEQNMISPNRSPSSSVREAKSYEGGY	435
QY	424 LEPRLGVHENIAVLDPSSMYNIMIKYNGSDTLVRGGEKCGECGCWEAPVKHRFRCP	483
DB	436 REPLKIQBDIVSLDPFMSLYPSILISHNISPTIY-----EKKER-ENME	480
QY	484 PGFFKTVLBRLELAKRYAEAMK---KVPDPSPEYRLIDBRQALKAVLANASVGMGS	539
DB	481 LGIIPKTNELLSRRKHIMLLKDKIQNEFD-EYSRLEHEQSIVKLANSHGYLAFF	539
QY	540 GARWYCBKCAVTAWRGHLIRTAINTIARKLGKVIYQDTSDFVTYDPKVE-----	592
DB	540 MARWYSDKCAEMVTGLGRKYIQETIEKAEERGFKVIADTDGFVAKWDYDKLQKKEEN	599
QY	593 -----NFKIIKESL--GFEIKLEKVKYR--LPFTEAKRYAGL	627
DB	600 DKSDKLSLNPPLKSKHEELIILTKFLKGINEELPGMELEFEGHFKRGLFTV--KKYALI	657
QY	628 LEDGRIDIVGEPAVNGDCELAKEVQTKVVEIVLTKSEVNKAVYVKIKVLEEGKVPI	687
DB	658 EDDGHLVKGLEVARDSNIAKTQQAIVINALLDEDGDVNLAKIKNTIDNLKKNIDK	717
QY	688 EKLIVNKTLSKLESYTTTEAPHVVAARMLSAGYRVSPGDKIGYVIVKGGGRISQAWPY	747
DB	718 NDLIIHTQLTKNIEBKSTAPHIEVAKKIKORGDSVRVGDIYSIIVKGRSRSISRA---774	
QY	748 FVWKDPSQIDVTYYVDHQIIPAAALRIILCYGITEKKLKAS	787
DB	775 ELLEVAGDYDINYIDNOVLPPVTRIMESIGISIDEELKNS	814

DE Endonuclease P1-PspI (BC 3.1.1.-.-) (Psp-GDB pol intein)].

GN POL.

OS Pyrococcus sp. (strain GB-D).

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus.

OK NCBI_TaxID=69013;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 493-517.

RX MEDLINE=94094330; PubMed=8269515;

RA Xu M.-Q., Southworth M.W., Merisha F.B., Hornstra L.J., Perler F.B.;

RT "In vitro protein splicing of purified precursor and the

RL identification of a branched intermediate.";

RT Cell 75:1371-1377(1993).

CC -!- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE

CC EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.

CC -!- FUNCTION: INTEIN ENCODED ENDONUCLEASES ARE THOUGHT TO MEDIATE

CC INTEIN MOBILITY BY SITE-SPECIFIC RECOMBINATION INITIATED BY

CC ENDONUCLEASE CLEAVAGE AT THE "HOMING SITE" IN GENE THAT LACK THE

CC INTEIN.

CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate

CC + {DNA} (N).

CC -!- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES

CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)

CC FOLLOWED BY PEPTIDE LIGATION.

CC -!- BIOTECHNOLOGY: Used in the PCR method because of its high

CC thermostability and low error rate. Sold by New England Biolabs.

CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.

CC -!- SIMILARITY: IN THE INTEIN SECTION; BELONGS TO THE HOMING

CC ENDONUCLEASE FAMILY.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@lsb-sib.ch).

CC -----

CC EMBL; U00707; AAA67130.1; -

CC EMBL; U00707; AAA67131.1; -

CC EMBL; U00707; AAA67132.1; -

CC PIR; S68593; S68593.

CC HSP; P56689; ITGO.

CC REBASE; 2619; PI-PspI.

CC InterPro; IPR006172; DNA_pol_B.

CC InterPro; IPR006134; DNA_pol_B_dom.

CC InterPro; IPR006133; DNA_pol_B_exo.

CC InterPro; IPR003586; Hedgehog_hintC.

CC InterPro; IPR003587; Hedgehog_hintN.

CC InterPro; IPR006141; Intein.

CC InterPro; IPR006142; Intein.

CC InterPro; IPR004042; Intein_endonuc.

CC InterPro; IPR004578; Pol2.

CC Pfam; PF00136; DNA_pol_B_2.

CC Pfam; PF03104; DNA_pol_B_exo; 1.

CC PRINTS; PR00106; DNAPOLB.

CC PRINTS; PR00379; INTEIN.

CC SMART; SM00305; HintC; 1.

CC SMART; SM00306; HintN; 1.

CC SMART; SM00486; POLBc; 1.

CC TIGRfams; TIGR01443; intein_Cterm; 1.

CC TIGRfams; TIGR01445; intein_Nterm; 1.

CC TIGRfams; TIGR00592; pol2; 1.

CC PROSITE; PS00116; DNA_POLYMERASE_B; 1.

CC PROSITE; PS50818; INTEIN_C_TER; 1.

CC PROSITE; PS50819; INTEIN_N_TER; 1.

CC PROSITE; PS50817; INTEIN_ENDONUCLEASE; 1.

CC Transferase; DNA-directed DNA polymerase; DNA replication;

CC DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;

CC Protein splicing; Intronic homing.

CC CHAIN 1 492 DNA POLYMERASE, 1ST PART.

CC CHAIN 493 1029 ENDONUCLEASE PI-PSPI.

CC CHAIN 1030 1312 DNA POLYMERASE, 2ND PART.

SQ SEQUENCE 1312 AA; 152852 MW; B62518805641D26A CRC64;

Query Match 23.1%; Score 965.5; DB 1; Length 1312;

Best Local Similarity 21.9%; Pred. No. 7.6e-53;

Matches 300; Conservative 154; Mismatches 282; Indels 635; Gaps 23;

QY 8 VLDSSYEVVCKEPOVINGIANGSERVVLIDRSFPYFYALLAPGADPKQV----AQRI 63

Db 2 ILDADYITDGGKPIIRFK-KENGDFKVEYDRNFRPYVALIKQDSQIDEVRKITAERHG 60

QY 64 ALSRPKSPFIIGVEDDKSKYFGPRVLRITVLPKAVREYRELKYNVGVGVDEADTRP 123

Db 61 KIVR-----IIDAEKVRKFLGRPIEWRWLYFEHPQDVPAIRDKISESAVIDIFSDIP 116

QY 124 AWRYLIDHDLFFPTWYRVEAEPLNKGFRVDKVLVKSREPEYLGALPTKLPDLRL 183

Db 117 AKRYLIDKGLI-----PMED-----EELKLL 138

QY 184 AFDIEVYSKQSPRPDPVIVIAKTDDGDEVL-----FIAGKDDRKPIRPFVEY 235

Db 139 AFDIETLYHEGE-EPAKGIIMISYADEEAKVITMKKIDLPYVEVSSEREMIKRFLKV 197

QY 236 VKRYDPDIIVGYNHNFDPVLLRRARILGKLDVTRVGAEPITSVHGH---VSVPRGL 292

Db 198 IREKDPDIITVYNGSFDLPVLVRAEKLGLKPLGRD-GSEPKQRIGDNTAVEIKRI 256

QY 293 NVLDYDYAEEMPEIKIKSLAEVAYGLVGMKKSERVIINWBEIPDYDDPKRPLLLQYAR 352

Db 257 HPDLYHVIRRTINLTPTYLEAVYEA1-FGKPKKYAH--EIAEAWETGKGLERVAKYSM 313

QY 353 DNVRTATYGLAEKLLPFAQLQSVTVGLPDLQVGMVSGRLEWYLIRAFKMKELVNRVE 412

Db 314 EDAKTYTBLGREFFPMEAOQLSRLVQQLVDVSRSTGNLVWYLLKAYERNELAPNKP 373

QY 413 RPE-----ETVGAIVLEPLRGVHENTAVLDSSMYENIMKYNVGDTLVRGKCGE 466

Db 374 REYERRLRRESVAGYVKEPEKGLWGLVSDLSFLPSIIITNVSPTLNR----- 427

QY 467 CGCWE---APEVGHFRRCPPGFFKTVLRLLELRKVRRAEMKKYPPDSPRYLLDRQK 523

Db 428 -GCREYDVAPEVGHFKCFKDFGFTPSLLKRLDERQEIQRKWA-SKQPIEKKMLDYQR 485

QY 524 ALKVL----- 529

Db 486 AIKILANSILPEEWVPLIONGKVKIFRIGDFVDGLMKANQGVKTKGTDEYLEVAGIHA 545

QY 530 ----- 529

Db 546 SFDRKSKKARVMKAVKAVIRHYSNVYRIVLNSGRKITYTEGHSIFVYRNGDLVETG 605

QY 530 ----- 529

Db 606 VKIGDLLAVPRSVNLPEKRLNIVELLNLNLSPEETEDILITIPVKGRKNFKGMLRTL 665

QY 530 ----- 529

Db 666 WIFGEKRVRTASRYLRHLENGLYRURKIGYDIIDKEGLEKVRTLYEKLVDVVRVYNGK 725

QY 530 ----- 529

Db 726 REYLVERNAVREDVLSLMBEELKEWRIGTRNGRMGTFFVIDEDFAKLGLGYVSEGSARK 785

QY 530 ----- 529

Db 786 WKNOTGGSVTYRLYNENDEVDLMEHLAKFFKGVKRGKYNVEIPKMAVITFESICGT 845

QY 530 ----- 529

Db 846 LAENKRVPEVITSSKGVKRWAFLEGYFIGDGVHPKRVLSTKSELLVNLGLVLLNSLG 905

QY 530 ----- 529

Db 906 VSAIKLGYDSGVYVYNEELKTEYRKKNVHSHVDPKDIKLTFTGKVFQKNISYKRP 965

QY 530 ----- 529
 Db 546 SFNRETSELSKVALIRHRYGKVIYIKLSGRRIKITSGLSHLSFVNRKGLVKVRGDE 605
 QY 530 ----- 529
 Db 606 LKPGDLVVPGLKPSKQVNLVLLKLPBETSIVMMIPVKGRKFPKGMKLTLY 665
 QY 530 ----- 529
 Db 666 WIFGGERPRTAGRYLKHRLRGYVVKLRKRGCEVLDWESLKRKYELTILKLNKYNKNS 725
 QY 530 ----- 529
 Db 726 RAYVVEFNSLRDVSLMPEIBELKEWIIIGEPGRPKIGTFIDVDDSFALILGYISSGVZEK 785
 QY 530 ----- 529
 Db 786 DRVKPHSKDQNVLEDIAKLAELKFKVRRGRGVTEVSGKISHAIFRVLAEGRKRIPEFT 845
 QY 530 ----- 529
 Db 846 SPMDIKVAPLXGLANGNABELTFSTYSKSELVNLQILLNSIGVSDIKIEHKGVRVYINK 905
 QY 530 ----- 529
 Db 906 KESSNGOIVLDSVESIEVEKYGVYVYDLSVEDNENFLVGLLYAHNSVYGYVYAKARW 965
 QY 544 YCRECAKAVTAWGPHLRTAINIARKLGLXVYCDTDSLFVY.---DREKVE.---NFI 595
 Db 966 YKRECAESVTAWGQYDLVRRLEARGFKVLVYDIDGLYATIPGVKDWEEVKRRALEFV 1025
 QY 596 KIIKEBLG--PEIKLEKVKYKLFTEAKRYAGLLEDGRIDIVGFAVRGDMCELAKYVQ 653
 Db 1026 DYINSKUPGVLESEVGFYARGFV--TKKYVALIDEGKIYVTKGLEIVARDWSEIANKETQ 1084
 QY 654 TKVVEIVUKTSVENKAVEYKVIKVELEBKVPKIEKLVIMKLSKLEETTRAPHVAA 713
 Db 1085 ARVLEALKHGNVBNKVIKVDVTKNTYEVPEKLVYEQITRINEYKAGHPHVA 1144
 QY 714 KRLMSAGVRVSPGDKIGVIVKGGGRISORAMPVYVVK--DP--SOLDIVYVDHDIIPA 769
 Db 1145 KRLMARGIKVAFGVIGVIVLGRGSPISKRA--ISIEBDFPKHKYDAEYIENQVLA 1201
 QY 770 ALRILGVFGITEKKLKASAGQKTLFDPL-AKUS 802
 Db 1202 VERILKARGYKREDLRWQTKQVGLGAWIKVKS 1235

RESULT 12

DPOL_METH
 ID DPOL_METH STANDARD; PRT; 586 AA.
 AC DP27276;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA polymerase (EC 2.7.7.7).
 GN POL OR POLB1 OR MTH1208.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Delta H;
 RX MedLine=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pochier B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;

RT
 RT detail: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
 CC + {DNA}(N);
 CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE000888; AB885697.1; -;
 DR PIR; C69028; C69028.
 DR HSP; P56689; ITGO.
 DR InterPro; IPR006172; DNA_pol_B.
 DR InterPro; IPR006134; DNA_pol_B_dom.
 DR InterPro; IPR006133; DNA_pol_B_exo.
 DR Pfam; PF001136; DNA_pol_B; 1.
 DR Pfam; PF03104; DNA_pol_B_exo; 1.
 DR PRINTS; PR00106; DNAPOB.
 DR SMART; SM00486; POLB; 1.
 DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
 KW Transferase; DNA-directed DNA polymerase; DNA replication;
 KW DNA-binding; Complete proteome.
 SQ SEQUENCE 586 AA; 67965 MW; D4065BC39B951A CRC64;
 Query Match 21.8%; Score 915; DB 1; Length 586;
 Best Local Similarity 35.4%; Pred. No. 3.9e-50;
 Matches 218; Conservative 110; Mismatches 226; Indels 62; Gaps 12;
 QY 1 MTEVVFVTLSSYEVVCKEPOVIINGAENG--ERVILDRSPRFPYFALLAPGADPQV 58
 Db 1 MEDYRVVLLDIDVTVDVEPVILFGKSKSGNEPIAHDRSPFYIAT-----PDLL 54
 QY 59 AQRTALSRKPSPIIGVEDDKRYGFRPRVRIRIIRTVLPEAVREVELKVNVDGVLE 118
 Db 55 DECLRELELELEKLEVE--MEDLGRPTVIRIEPRHPQDPVKIRDRIDLESVRDRE 112
 QY 119 ADYFAMRYLIDHOLF-----FTWYVRAEAPLENKMGPRVKVYVSRPPLVGEALA 173
 Db 113 HDIPFVRYLYIDKSIVPMEELEFGVGVDSAPSTVTDVETVETGRVQSTGSGNHG-- 168
 QY 174 PTKLPDLIRILAFDIEVYKSGSPRPRDPVIVIAVK-----TDDGVLFIAEGKD 224
 Db 169 -----LDILSFDIERNPHGMDPEKDEIVMIGVAGNMGYESVISTAGHLDPVVVED 222
 QY 225 DRKPIREFEYVKRYDDIIVGYNHFDWPHYLLRRARILGLKV-----TRVGA 276
 Db 223 ERELLERFAEIVIDKSPDLVNGYNSDNDFPYITRRAAILGAELDGLMGDSKIRTMKRGF 282
 QY 277 EPTTSVGHVSVPGRNLVDLYDAEEMPEIKLSLEAEVYGLVGMKKSRIINWEIPD 336
 Db 283 ANATAIKGV-----HVDLYPVWRVYMNLDRTVLEVVQELFGEKIDLPGRDNE--- 333
 QY 337 YWDDPKRPLLLQYARDVRAVYGLAEKILPFAIQLSYVYTGLPDQVGMVSGFLWEYL 396
 Db 334 YWRDELDLFRYSLDDVAVTHIAEKILPLNLELTRLVGQPLFDISMATGQQAWEFL 393
 QY 397 IRAAFKMKELVPRVERPETVE-----GAVLSEPLRGVHENIAVLDPSSMYNIMIKY 450
 Db 394 VRKAIQYGLVPMKPSQSFSSRRGRVAGVYKEPEKGLHNI VQDFRSLYPSIIISK 453
 QY 451 NVGPDTLVRPGEKCGCGWEAFVXKHRRFCPPGFFKTVLRLLELRKRVAEAKKYP 510
 Db 454 NISPDLTDDDES-----ECVAPEYGVFRKSPRGVPSVIGKILSERVRIKEEMKG-SD 508
 QY 511 DSEYRLDDEKRAIKVLANASVGMVSGARWYCECAKATVAGRHILRTAINIARKL 570
 Db 509 DPMERKILNVQQAALKRLANTMYGVYGRFRWYSGMECAEATVANGDYIKTKITKABEF 568

QY 571 GLKVIYGDTSLSFVY 586
 Db 569 GFHTVYADTGATY 584

RESULT 13

DPOL THERM STANDARD; PRT; 1523 AA.
 AC P74918;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA polymerase (EC 2.7.7.7) (Pol Tfu) [Contains: Endonuclease PI-TfuI
 (EC 3.1.-.-) (Tfu pol-1 intein); Endonuclease PI-TfuII (EC 3.1.-.-)
 (Tfu pol-2 intein)].
 GN POL.
 OS Thermococcus fomicolans.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Thermococcus.
 OX NCBI_TaxID=46540;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SF557;
 RA Camdon M., Querellou J.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP CHARACTERIZATION OF INTEINS.
 RC STRAIN=SF557;
 RX MEDLINE=20112788; PubMed=10644683;
 RA Saves I., Oxanne V., Dietrich J., Masson J.-M.;
 RT "Inteins of Thermococcus fomicolans DNA polymerase are endonucleases
 with distinct enzymatic behaviors";
 RL J. Biol. Chem. 275:2335-2341(2000).
 CC -!- FUNCTION: PI-TfuI recognizes and cleaves a minimal sequence of 16
 base pairs (bp) on supercoiled DNA with either Mn(2+) or Mg(2+) as
 cofactor. It cleaves linear DNA only with Mn(2+) and requires a
 19-bp minimal recognition sequence. The optimal temperature for
 activity is 70 degrees Celsius.
 CC -!- FUNCTION: PI-TfuII is a highly active homing endonuclease using
 Mg(2+) as cofactor. Its minimal recognition and cleavage site is
 21 bp long either on linear or circular DNA substrates. Its
 endonuclease activity is strongly inhibited by the 3' digestion
 product, which remains bound to the enzyme after the cleavage
 reaction. The optimal temperature for activity is 70 degrees
 Celsius.
 CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
 + {DNA}[N].
 CC -!- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
 A POST-TRANSLATIONAL EXCISION OF THE TWO INTERVENING REGION
 (INTEINS) FOLLOWED BY PEPTIDE LIGATION.
 CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
 CC -!- SIMILARITY: IN THE INTEIN SECTION; BELONGS TO THE HOMING
 ENDONUCLEASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; Z69882; CAA33738.1; -;
 DR HSP; P56689; ITGO.
 DR REBASE; 4500; PI-TfuI.
 DR REBASE; 4501; PI-TfuII.
 DR InterPro; IPR006172; DNA pol B.
 DR InterPro; IPR006134; DNA pol B dom.
 DR InterPro; IPR006133; DNA pol B-exo.
 DR InterPro; IPR003586; Hedgehog hintc.
 DR InterPro; IPR003587; Hedgehog hintn.
 DR InterPro; IPR006141; InteIn.

DR InterPro; IPR006142; INTEIN
 DR InterPro; IPR004042; InteIn_endonuc.
 DR InterPro; IPR004578; Pol2.
 DR Pfam; PF00136; DNA_pol_B; 2.
 DR Pfam; PF03104; DNA_pol_B_exo; 1.
 DR PRINTS; PR00379; INTEIN.
 DR SMART; SM00305; Hintc; 2.
 DR SMART; SM00306; Hintn; 2.
 DR SMART; SM00486; PolBc; 1.
 DR TIGRFAMs; TIGR01443; intein_Cterm; 2.
 DR TIGRFAMs; TIGR01445; intein_Nterm; 2.
 DR TIGRFAMs; TIGR00592; pol2; 1.
 DR PROSITE; PS00116; DNA POLYMERASE B; FALSE_NEG.
 DR PROSITE; PS00818; INTEIN_C_TER; 2.
 DR PROSITE; PS00819; INTEIN_ENDONUCLEASE; 2.
 DR PROSITE; PS00817; INTEIN_N_TER; 2.
 KW Transferase; DNA-directed DNA polymerase; DNA replication;
 KW DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;
 KW Protein splicing; Intronic homing; Magnesium; Manganese.
 FT CHAIN 1 406 DNA POLYMERASE, 1ST PART.
 FT CHAIN 407 766 ENDONUCLEASE PI-TFU I.
 FT CHAIN 767 900 DNA POLYMERASE, 2ND PART.
 FT CHAIN 901 1282 ENDONUCLEASE PI-TFU II.
 FT CHAIN 1283 1523 DNA POLYMERASE, 3RD PART.
 SQ SEQUENCE 1523 AA; 175917 MW; 7A2AC823BF25F5 CRC64;
 Query Match 20.2%; Score 845; DB 1; Length 1523;
 Best Local Similarity 18.9%; Pred. No. 3,4e-45;
 Matches 297; Conservative 149; Mismatches 299; Indels 828; Gaps 23;
 QY 8 VLDASVEYVGKPEPVIITWGIANGSERVVLIDRSRPYFYALLAGADPKQVAQRILASR 67
 Db 2 ILDTYITDGRPVIRVFK-KENGFEKIEYDRDPFYIALLKODSAIEDVKKITASRHG 60
 QY 68 PKSPITIGVEDDKRYFGSRPRVLRTITVLPFAVREYBELVKNVGVGVEDVLEADIREAMRY 127
 Db 61 TTVVRVAGKVKKFLGSPVIEVWKLTHYPODVPAIRDKREHPAVVDIYEDYIPFARKY 120
 QY 128 LIDHDLFPFTVYRVEABPLENKGFRVDKVLVYKSRPELYGEALPTKLPDLRLAFDI 187
 Db 121 LIDKGLI-----PMEGD-----BELKWLAFDI 142
 QY 188 EYVSKQSGPRPRDPVIVIAVKTDDGDEVL-----FIAEGKDDKAPREFVEYVKRY 239
 Db 143 ETLHGEFEAE-GPILMISVADEEGARVITWKIDLPYDVVSTKEMIKRFLKVVYKEK 201
 QY 240 DPDIIVGNHNPWPYLLRARIILGKLDVTRVGAEPSTTVHGH---VSVQRLNVDL 296
 Db 202 DFDVLITVNGDNFDFAVKKSEKLGVKF-ILGRDSEPKIQMGDRFAVEVKGRIHFDL 260
 QY 297 YDAEEMPEIKISLEVAEVLGVNKKSERVINNNWIPDYDDPKRPLLOVARDVVR 356
 Db 261 YPVIHNTINLPTVYLEAVYEAIFGQPK-EKYVAB--ETAAQWETGEGLERVARVSMEDAK 317
 QY 357 ATYGAEKILPFAIQLSVVTGLPDDQVGMVGRLEWYLIRAFPKMKELVPR-----V 411
 Db 318 VTSELGFEFFPMERQSLVGVQSPWVDSRSTGNLWVYLLRKYAERNEINLPKPSREL 377
 QY 412 ERPBETVGAIVLPLRGVHNIIVLDF----- 439
 Db 378 EERGGVAGGVKVEPERGLWENIAYLDFRCHPADTKVIVKGVGVNISEVRREGDVLGID 437
 QY 440 ----- 439
 Db 438 GWQKVQRWVEYDEGLVNLNGLKCTPNHKLPVVRTERQTATRDLSLAKSLTKVKVKGKL 497
 QY 440 ----- 439
 Db 498 ITTPLFEKIGKIEREDVPEEILKGLAGAILAEGTLRLKDKVEYFDSRGKRVSHQYVR 557
 QY 440 ----- 439
 Db 558 EITVGAQEEDEFQRIIVYIFELFGVTFPSVYRKQNTNATITFKVAKKEVYLVRIMDGIEN 617

QY 440 ----- 439
 Db 618 LHAPSVLRFPGDGSVNVKRTVVVQGTNNENKIEWWSKLNKLGIPHRRVYDYTER 677
 QY 440 ----- 439
 Db 678 EKTMTTHLEIAGRDGLILFQITVGFISTERNKMALEEAIRNRVNLNNAFYTLADFTA 737
 QY 440 ----- 471
 Db 738 KTEYKGVYDITLBTGTPYFANGILTHNSLYPSLISHNVSPDTLNRG--CGEYD--E 793
 QY 472 APEVXHPRCPPGPFKTVLEKLELRKVRRAEMKYPDPDPSEYVLLDDEKQAKVLVANA 531
 Db 794 APQVGHFRFCDFGPFIPSLGDLDERQKVKHKA-TVDPTEKLLDYRQEAITILANS 852
 QY 532 SYGYMGSGARVYCEKAKAVTAWGRHLIRTAI-NIAREKLGKVY-----GDTD----- 580
 Db 853 FYGYGAKARVYCEKAESEVTAWGRQYIETTRIEBKFGKVLVADSVTGDTEVTRR 912
 QY 581 ----- 580
 Db 913 NGRIEFVPIEKLPERVDRHVRGEKVCYVGLGVEALFLDNRGLVWVKVYVVRHKTDRY 972
 QY 581 ----- 580
 Db 973 RVNFTNSWYLDVTEHSLGYLNTSKVPGKPLKRLVYVKEPBLGGKVKSLTPNPIA 1032
 QY 581 ----- 580
 Db 1033 RTIKANPIAVKLWELIGLLVGDNGWQSNWAKYVYVGLSCGLDKAEIERKVLNPLREASV 1092
 QY 581 ----- 580
 Db 1093 ISNYDKSKGQVSLSKWLAFGMVYKFDENGKAIKPSFMNLPREYIEAFLEGLPSAD 1152
 QY 581 ----- 583
 Db 1153 GTVSLRGRGPIEIRLTSVNRSLDAVRKULLVLGVNSLFTETKPNRLEKESGTHSR 1212
 QY 584 ----- 587
 Db 1213 IKNHRFADRIGLDRKSTKLSNGLGHTNKRAKYKDFDLVTPRKIESTYDGYVDI 1272
 QY 588 ----- 612
 Db 1273 EYEGTHRFANGILVHNTDGFATIPGADAETVKKKAREFLNYINPKLPGLLELEYSGFY 1332
 QY 613 KRLFFTEAKRYAGLLEDGRIDIVGFEAVRGDMCLAKEVQTKVVEIVLKTSEYVNAVEY 672
 Db 1333 RRGFFV-TKKYVAVIDEBGKITRGLIIVREDMSVAKETQARVLEAILRHGVDSEAVRI 1391
 QY 673 VRKIVKELEGKVPTEKLVIMTKLSRLEEYTTAPHVVAARMLASGVYRSPGDKTYV 732
 Db 1392 VKEVTEKLSYEVPEKLVIHQITRELKQYKATGPHVAIAKRLAARGIKVPTVLSYI 1451
 QY 733 IVKGGGRISQRAWPYFMWADPS--QIDVTYVVDHQIIPAALRILGYFGITEKKLKASATG 790
 Db 1452 VLKSGRIGDRTIP-FDEFDPTKRVDAEYIENQVLPFAVERILLKAFGYKKEDLRYQKTR 1510
 QY 791 QKTLEDFLAKKK 803
 Db 1511 QVGLGAWLKWGKK 1523

RESULT 14

ID DPOL_PYRKO
 AC DPOL_PYRKO STANDARD; PRT; 1671 AA.
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE DNA polymerase (EC 2.7.7.7) [Contains: Endonuclease PI-Pkoi
 DE (EC 3.1.-.-) (pko pol-1 intein) (IVS-A); Endonuclease PI-Pkoi
 DE (EC 3.1.-.-) (pko pol-2 intein) (IVS-B)].
 GN POL.
 OS Pyrococcus kodakaraensis.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Thermococcus.
 OX NCBI_TaxID=69014;
 RN [1] SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RP STRAIN=KOD1;
 RC MEDLINE=98027387; PubMed=9361436;
 RX Takagi M., Nishio M., Kakiyama H., Kitabayashi M., Inoue H.,
 RA Kawakami B., Oka M., Imanaka T.;
 RA "Characterization of DNA polymerase from Pyrococcus sp. strain KOD1
 RT and its application to PCR.";
 RT Appl. Environ. Microbiol. 63:4504-4510 (1997).
 RL [2]
 RN CHARACTERIZATION OF INTEINS.
 RP STRAIN=KOD1;
 RC MEDLINE=98416199; PubMed=9742242;
 RX Nishio M., Fujiwara S., Takagi M., Imanaka T.;
 RA "Characterization of two intein homing endonucleases encoded in the
 RT DNA polymerase gene of Pyrococcus kodakaraensis strain KOD1.";
 RL Nucleic Acids Res. 26:4409-4412 (1998).
 CC -1- FUNCTION: INTEIN ENCODED ENDONUCLEASES ARE THOUGHT TO MEDIATE
 CC INTEIN MOBILITY BY SITE-SPECIFIC RECOMBINATION INITIATED BY
 CC ENDONUCLEASE CLEAVAGE AT THE "HOMING SITE" IN GENE THAT LACK THE
 CC INTEIN. PI-PKOI RECOGNIZES 5'-GATTAGTCCCTGACC-3' AND PI-PKOII
 CC RECOGNIZES 5'-CAGCTACTACGGTAC-3'. BOTH ARE THERMOSTABLE.
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
 CC + [DNA] (N).
 CC -1- P.TM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
 CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
 CC FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
 CC -1- SIMILARITY: IN THE INTEIN SECTION; BELONGS TO THE HOMING
 CC ENDONUCLEASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D29671; BAB06142.2; -
 CC EDB; 1GCK; 28-JAN-03.
 CC REBASE; 3792; PI-PKOI.
 CC REBASE; 3793; PI-PKOII.
 CC InterPro; IPR006172; DNA_pol_B.
 CC InterPro; IPR006134; DNA_pol_B.
 CC InterPro; IPR006133; DNA_pol_B.
 CC InterPro; IPR003586; Hedgehog_hintc.
 CC InterPro; IPR003587; Hedgehog_hintc.
 CC InterPro; IPR006141; intein.
 CC InterPro; IPR006142; intein.
 CC InterPro; IPR004042; intein_endonuc.
 CC InterPro; IPR004578; pol2.
 CC Pfam; PF00136; DNA_pol_B.
 CC Pfam; PF03104; DNA_pol_B.
 CC PRINTS; P00379; INTEIN.
 CC SMART; SM00305; Hintc; 2.
 CC SMART; SM00306; Hintc; 2.
 CC SMART; SM00486; POLB; 1.
 CC TIGRfam; TIGR01443; intein_Cterm; 2.
 CC TIGRfam; TIGR01445; intein_Nterm; 2.
 CC TIGRfam; TIGR00592; pol2; 2.
 CC PROSITE; PS00116; DNA_POLYMERASE_B; 1.
 CC PROSITE; PS00818; INTEIN_C_TER; 2.
 CC PROSITE; PS00819; INTEIN_ENDONUCLEASE; 2.
 CC PROSITE; PS00817; INTEIN_N_TER; 2.

KW Transferase; DNA-directed DNA polymerase; DNA replication;
 KW DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;
 KW Protein splicing; Intron homing; 3D-structure.

FT CHAIN 1 406 DNA POLYMERASE, 1ST PART.
 FT CHAIN 407 766 ENDONUCLEASE PI-PKO I.
 FT CHAIN 767 851 DNA POLYMERASE, 2ND PART.
 FT CHAIN 852 1388 ENDONUCLEASE PI-PKO II.
 FT CHAIN 1389 1671 DNA POLYMERASE, 3RD PART.
 SQ SEQUENCE 1671 AA; 193490 MW; 4A1F1C8120EE455 CRC64;

Query Match 18.5%; Score 776; DB 1; Length 1671;
 Best Local Similarity 17.6%; Pred. No. 8.2e-41;
 Matches 304; Conservative 145; Mismatches 287; Indels 990; Gaps 24;

```

QY 8 VLSSSEVVGKSPQVINGTAENGVRVLLDRSFRYPFYALLAPGADPKQV-----AQRIR 63
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2 ILSDTYTIDGKGVIRIFK-KENGFEKIEYDRTFFPYFALLKDDSAIBEKVKITAERHG 60
QY 64 ALSRPSPIIGVEDDKKYGRRRLRLRTVLPEAVREYRELKVNVDGVEDLEADIRP 123
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 TVTVTKR-----VBKVKKFLGRPVWKLXPTHQDVPALRDKIRHPAVIDIYEDIRP 116
QY 124 AMRYLIDHDLPFTTWYRVEAEPLKNGKFRVDRVYLKSRPEPLYGEALAPTLPDLRL 183
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
117 AKRYLIDKGLV-----PMEGD-----BELKML 138
QY 184 AFDIEVYSKGSRRPRDPVIVIAVKTDDGDEVL-----FIAGKDDRPPIREFVEY 235
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
139 AFDIETLYHSGEFAB-GPTLMISYADEGARVITWKNVDLPYDVVYSTEREMIKRFLRV 197
QY 236 VKRYPDPIIVGYNHFDMPYLLARRILGILKDVTRRVGABPTTSVHGH---VSPGRL 292
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
198 VKKQDPDLTYGDNFDFAYLKKRCEKLGINFALGRD-GSEPKIQMGDRFAVEVKGR 256
QY 293 NVLDYDAERMPRIKLSLEVAEYLGVMKKSERVINWMEIPDYWDPKRPLLLQYAR 352
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
257 HFDLYPVIRRTINLPTVLEAVYAVFGQPK-EKVYAE-EITAMETGENLSEVARYSM 313
QY 353 DDVRAITYGLAEKLPFAIQLSYTGPLDQVGMSVGRLEWLIIRAAFMKELVNRYE 412
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
314 EDKAVTYELKEFLPMEQAQSLRIGQSLDWSRSSGNLVEFWFLLRKAVERNLAPNKP 373
QY 413 RPE-----ETRYGAIVLEPLRGVHENIAVLDF-----439
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
374 EKELARRQSYEGYKVEPGRGLENIVILDFRCHPADTKVYVKGKGIINISEVQSGDYV 433
QY 440 -----439
DB 434 LGIDGMQRVKWWEYDKGBLVNGLKCTPNHKLPPVTKNERQTRIRDSLAKSPLTKKV 493
QY 440 -----439
DB 494 KGKIITPLFYEGRATSENIPEEVLKGBLAGILLAEGLTLRKQVVEYFDSSRKKRISH 553
QY 440 -----439
DB 554 QYRVEITGKDEBEFRDRITYIERLFGITPSISEKGTNAVTLKVAKNVYLKVKELMD 613
QY 440 -----439
DB 614 NIESSLHAPSVLGRFFEGDGSVNRVRSIVATQGTQKNEWKILVSKLLSQLGIPHQTYVQ 673
QY 440 -----439
DB 674 YQENGKORSYILEITGKGLILFQTLIGFISERKNALLNKAISQREMNLENNGFYRLS 733
QY 440 -----467
DB 734 EFNVSTYEGKVDLTLEGTTPYFANGILTHSLPSIIITNVSPDLNRE-----786
QY 468 GCWGE---APEYKGRFRCPFGFFKTVILERLELKRVRAEKPKYPPDSEYRLDRQKA 524
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
787 GCXEYDVAPQVGHFRKDFGFTPLSLGDLLEERQKIKKQKA-TIDPIERKLLDYRQRA 845

```

```

QY 525 LKVLK-----529
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
846 IKILANSILPBEWLPVLEESGVHFRIGELIDRMMEENAGVKRGEGETEVLEVSGLVEFS 905
QY 530 -----529
DB 906 FNRRTKAEKLRVKALIRHDSGKVVYTIKLSGRRIKITSGLSHLSFSVNGELVEVTGDEL 965
QY 530 -----529
DB 966 KPGDLVAVPRLELPERNHVLNVELLIGTPEETLDIVMTIPVKGKKNFFKGLMLTLRW 1025
QY 530 -----529
DB 1026 IFGEKRRPRTARRYLRHLEDLGYYRLKKGIVELDWDLSLKNRYRLYEALVENRYNGKR 1085
QY 530 -----529
DB 1086 EYLVEFNSIRDVAGIMPLKELKWKIGTLNGFRMKLIEVDESLSKLLGYVYVSEGYARKQ 1145
QY 530 -----529
DB 1146 RNPKNWSYSKLYNEDPEVLDMDMERLASRFFGKVRGRNYVEIPKIGYLLFFENMCGVL 1205
QY 530 -----529
DB 1206 AENKRIPEFVFTSPKGVRLAFLEGYPTGDDVHNKRLRLSTKSELLANQVLVLLNSGV 1265
QY 530 -----529
DB 1266 SAVKLGHDGCVYRVYINEELPFVKLDKKKNAYSHVIPKVELSEVFGVKQKNSVQPIFR 1325
QY 530 -----529
DB 1326 KMVEDGRLDPEKAQRLSWLIEGDVVLDREVSDVDYGVYDLSVEDNENFLVGFGLVY 1385
QY 530 -----586
DB 586 --NAGSYGMGSGARWYCRECAKAVTANGRLIRTAI-NIARKLGLKVIYGDTSLSFVY 586
QY 1386 ANSYGYGYGARAWYCKEASVTANGREYIWTIKIEEKGKGFYIYSDTGFFAYI 1445
QY 587 ---DPERVE---NFIKIKBEL--GFEIKLEKVKRLFFTEAKKRYAGLLEDGRIDIVG 637
DB 1446 PGADAETVKKAMEFLKVINAKLPGALEYEGFYKRGFFV-TKKYAVIDEKGIITRG 1504
QY 638 FEAVRGDCWELAKEVQTKVVELVKTSEVNAKAVEYRKIVKELBEGKVPDKLVIWKTLS 697
DB 1505 LEIVRDRWSEIAKETQAEVLEALLKDGQVEKAVRIVKSEVTKLSKYEVPEKLVTHEQIT 1564
QY 698 KLEBEYTTAEAPHVAAKMLSAGYRVSPDGKIGVIVKGGGRISORAMPYFWKDPDS--Q 755
DB 1565 RDLKDYKATGPHAVAKMLAARGVKIRGTVISIYLVKSGSRIGDRAIP-FDEFDPTXHK 1623
QY 756 IDVTYVVDHQIIPAAIRILGYFITEKKLKSATGQKTLFDFLAKK 801
DB 1624 YDAEYTIENQVLPAVERILRAFGYRKEDLRVQKTRQVGLSAWLKPK 1669

```

RESULT 15

DPOD SCHPO

ID DPOD SCHPO STANDARD; PRT; 1086 AA.

AC P30316; Q10016; Q9USU0; Q9UU61;

DT 01-APR-1993 (Rel. 25, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE DNA polymerase delta catalytic subunit (EC 2.7.7.7) (DNA polymerase

DE III).

GN POL3 OR POLD OR SPBC336.04.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI_TaxID=4896;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 14:56:08 ; Search time 42 Seconds
(without alignments)
4933.720 Million cell updates/sec

Title: US-10-034-849-2
Perfect score: 4188
Sequence: 1 MTEVFTVLDSSYEVVGKEP.....LKASATGQKTLDFLAKSK 803

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mmc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3033.5	72.4	803	1 Q59691	Q59691 pyrodicticum
2	2248	53.7	781	17 Q96YV1	Q96YV1 sulfobolus
3	2246	53.6	781	1 Q9P9N1	Q9P9N1 sulfurispha
4	2156	51.5	785	17 Q8ZVU2	Q8ZVU2 pyrobaculum
5	2129.5	50.8	785	1 Q9P9M2	Q9P9M2 pyrobaculum
6	1268	30.3	775	1 Q9P9K4	Q9P9K4 pyrococcus
7	1264	30.2	775	1 Q9HH06	Q9HH06 pyrococcus
8	1208	28.8	759	1 Q9HH98	Q9HH98 methanopyru
9	855	20.4	933	17 Q8PVC1	Q8PVC1 methanosarc
10	852.5	20.4	937	17 Q8TSB3	Q8TSB3 methanosarc
11	760.5	18.2	1107	4 Q9EH98	Q9EH98 homo sapien
12	760.5	18.2	1107	4 Q9ENR3	Q9ENR3 homo sapien
13	744.5	17.8	1105	11 Q8C2N0	Q8C2N0 mus musculu
14	743.5	17.8	1105	11 Q91VT0	Q91VT0 mus musculu
15	737.5	17.6	1104	3 Q8X0N7	Q8X0N7 neurospora

17	732.5	17.5	974	5	Q8SQF5	Q8SQF5 encephalito
18	718.5	17.2	901	17	Q9HRV9	Q9HRV9 halobacteri
19	655.5	15.7	855	17	Q8ZVP9	Q8ZVP9 pyrobaculum
20	649.5	15.5	914	1	Q59690	Q59690 pyrodicticum
21	648	15.5	876	17	Q971C7	Q971C7 sulfobolus
22	646.5	15.4	796	17	Q9HJRO	Q9HJRO thermoplas
23	622	14.9	800	17	Q97AH3	Q97AH3 thermoplas
24	575.5	13.7	1458	13	Q9DE46	Q9DE46 xenopus lae
25	569	13.6	787	16	Q9KSP4	Q9KSP4 vibrio chol
26	564.5	13.5	844	1	Q31096	Q31096 cenarchaeum
27	560.5	13.4	1016	12	Q993K6	Q993K6 callitrichi
28	552	13.2	845	1	O74046	O74046 cenarchaeum
29	540	12.9	1492	10	Q9FHA3	Q9FHA3 arabidopsis
30	539	12.9	2909	4	Q9NU25	Q9NU25 homo sapien
31	536	12.8	1015	12	Q8UZD7	Q8UZD7 cercopithic
32	532.5	12.7	1488	5	Q9VD90	Q9VD90 drosophila
33	527.5	12.6	787	16	Q91211	Q91211 pseudomonas
34	524	12.5	1013	12	Q91IX9	Q91IX9 retroperito
35	521.5	12.5	1013	12	O41208	O41208 retroperito
36	520	12.4	808	16	Q8EFZ4	Q8EFZ4 shewanella
37	515	12.3	1048	12	Q69390	Q69390 pseudorabie
38	512	12.2	789	16	Q8D8C5	Q8D8C5 vibrio vuln
39	511	12.2	956	5	Q8SRW2	Q8SRW2 encephalito
40	511	12.2	1001	12	Q8AYV2	Q8AYV2 porcine lym
41	511	12.2	1026	12	O36363	O36363 alcelaphine
42	510	12.2	1004	12	Q9Q6Z8	Q9Q6Z8 porcine lym
43	507.5	12.1	1428	5	Q9NAH1	Q9NAH1 caenorhabdi
44	506.5	12.1	1012	12	O40915	O40915 kaposi's sa
45	506.5	12.1	1012	12	O40910	O40910 kaposi's sa

ALIGNMENTS

RESULT 1

Q59691 ID Q59691 PRELIMINARY; PRT; 803 AA.
AC Q59691:
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN POLB.
OS Pyrodicticum occultum.
OC Archaea: Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Pyrodictiaceae; Pyrodicticum.
OX NCBI_TaxID=2309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM2709;
RX MEDLINE=95238290; PubMed=7721707;
RA Temori T., Ishino Y., Doi H., Kato I.;
RT "The hyperthermophilic archaeon Pyrodicticum occultum has two alpha-like DNA polymerases.";
RL J. Bacteriol. 177:2164-2177 (1995).
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE + {DNA}(N).
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
DR EMBL; D88574; BA07580.1; -.
DR HSSP; P56689; ITGO.
DR InterPro; IPR006172; DNA pol B.
DR InterPro; IPR006134; DNA pol B dom.
DR InterPro; IPR006133; DNA pol B_exo.
DR InterPro; IPR004578; Pol2.
DR Pfam; PF00136; DNA pol B; 1.
DR Pfam; PF03104; DNA pol B_exo; 1.
DR SMART; SM00486; POLBc; 1.
DR TIGRfam; TIGR00592; pol2; 1.
DR PROSITE; PS00116; DNA POLYMERASE B; 1.
KW DNA replication; DNA-binding; DNA-directed DNA polymerase;
KW Nucleotidyltransferase; transferase.
SQ SEQUENCE 803 AA; 92656 MW; E9C092F26A8D23FA CRC64;


```

Db 589 EIKDKIKYKRVFFTENKKRYAGLTEDGKIDIVGFEAVRGDWCDLAKQVQNTNVLILKSG 648
Qy 665 EVNKAVEVVRKIVKELEBGKYPPEKLVITWKTLSKLEBYTTEAPHVVAARMLSAGYRVS 724
Db 649 KVEDAIKVKVIVPDRYRNFIEDLIITWKTIDQNDDEYDVTAPHVVAARMLSAGYRVS 708
Qy 725 PDGKTGYVIVKGGGRISORAMPYFMVDPDSQIDVYVVDHQIIPAAIRILGYFGEITEKKL 784
Db 709 KGKIGYVIVKGGGRISORAMPYFMVDPDSQIDVYVVDHQIIPAAIRILGYFGEITEKKL 784
Qy 785 KASATGQKTLDFDLAK 800
Db 769 K---TGGVDILSFFKK 781

RESULT 3
ID Q9P9N1 PRELIMINARY; PRT; 781 AA.
AC Q9P9N1;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DE 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE DNA polymerase B3 (EC 2.7.7.7).
OS Sulfurisphaera ohwakensis.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfurisphaera.
OX NCBI_TaxID=69656;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TA-1;
RX MEDLINE=20450680; PubMed=10997874;
RA Iwai T., Kurosawa N., Itoh Y.H., Kimura N., Horiuchi T.;
RT "Sequence analysis of three family B DNA polymerases from the
thermoacidophilic crenarchaeon Sulfurisphaera ohwakensis.";
RL DNA Res. 7:243-251(2000).
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE
+ [DNA](N).
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
DR EMBL; AB032376; BAA93703.1; -.
DR HSP; P56689; ITGO.
DR InterPro; IPR006172; DNA pol B.
DR InterPro; IPR006134; DNA pol B dom.
DR InterPro; IPR006133; DNA pol B_exo.
DR InterPro; IPR004578; Pol2.
DR Pfam; PF00136; DNA pol B; 1.
DR Pfam; PF03104; DNA pol B_exo; 1.
DR SMART; SM00486; POLB; 1.
DR TIGRFAMs; TIGR00592; pol2; 1.
DR PROSITE; PS00116; DNA POLYMERASE B; 1.
KW DNA replication; DNA-binding; DNA-directed DNA polymerase;
KW Transferase.
SQ SEQUENCE 781 AA; 90384 MW; 9B2570EA30C372C CRC64;

Query Match 53.6%; Score 2246; DB 1; Length 781;
Best Local Similarity 54.3%; Pred. No. 4.5e-146;
Matches 432; Conserved 141; Mismatches 203; Indels 20; Gaps 10;

Qy 6 FTVLDSYVVKQKQVQVINGAENGVRVLLIDRSFRFYFALLAPGADPKVQVQVIRAL 65
Db 5 FFLDPSYVNVNKPVIYIWIWDKGNRYVLLKXFRFYFVYVYVYVYVYVYVYVYVYVYV 64
Qy 66 SRPKSLIGVEDKRYKGRPRVLRIRTVLPEAVREYRELKVNVDGVEDVLEADIRFAM 125
Db 65 SKPTSTIDVEEKIFGSPVKVLKIEVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 124
Qy 126 RYLIDHDLFFFTYRVEAPELNGKGRVDRKVLVYKSRPEPIYGEALPTKLPDLIRLAF 185
Db 125 RYSDINLAFYVWIEAEVEEIKEN-NFRVKVYELK-KINKLYED----KIPSLKVLAF 177
Qy 186 DIBVSKQSGPRPRDPVIVIAVKTDDGDEVLFIAEGKDDKRPREFVEYVKRYPDILV 245

```

```

Db 178 DIEVNYKGSNPRDRPVIIIGVWTKEGGK-QFLADKYDOLRAIRBFINFVQYDPDIIV 236
Qy 246 GYNNHHPWYLLRARIIGIKLDVTRRVGAETTSVHGHSVPGRLNVDLYDIABEMPE 305
Db 237 GYNNHHPWYLLRARIIGIRLDVGRVNGEPEQGVYHSITGRNLVDLYGFAQSIQE 296
Qy 306 IKISLEBVAIYLGWKKSERVINWWEIPDYWDPKRFLLOIYAADDDVRAIYGLAEKI 365
Db 297 VKVKTLENIADYGLPKERTIVEMDYIPKYWDDKXKRDILLKYNLDLAKSAYLLGEVF 356
Qy 366 LPFAIQSYVYVGLDQVGAMSVGRLEWLVIRAAFPKMKELVNRVERPETYRGAIVLE 425
Db 357 IPFGIEUTRISGLFDQLSMASVGRVWELLKREAYKYNELIPKGEREYESYEGGLVIS 416
Qy 426 PLRGVHENIAVLDFSSMYPNIMIKYVGPDTLVRPGEKCGEC-GCWEAPKVRHFRRCPP 484
Db 417 PLPGIHEEVVLDFFSSMYPSIMIKYVGPDTLVK----GECENCWVSP-VGHKFRKEPP 470
Qy 485 GFFKTVLERLELRKRYPAEMKKYPPDSPYVLLDEROKALKVLANASYGVMSGARWY 544
Db 471 GLYKNVLEKLTQERREVKKMEK-TIDEYDKRVLDAQRALKVNNANAFYGMGLGARWY 529
Qy 545 CRECAKAVTANGRLHIRTAINAEKLGKLVYIGTDSLFTYDPEKVENFIKIKSELGF 604
Db 530 SKEGAEVATANGRQIISDSAKIAKEKGTVIYIGTDSIFVKGGSD-INSLITEISSKFL 588
Qy 605 EIKLKVYKRLFFTEAKRYAGLLEDGRIDIVGFEAVRGDWCDLAKQVQNTNVLILKSG 664
Db 589 EIKIDIKYKRVFFTENKKRYAGLTEDGKIDIVGFEAVRGDWCDLAKQVQNTNVLILKSG 648
Qy 665 EVNKAVEVVRKIVKELEBGKYPPEKLVITWKTLSKLEBYTTEAPHVVAARMLSAGYRVS 724
Db 649 KVEDAIKVKVIVPDRYRNFIEDLIITWKTIDQNDDEYDVTAPHVVAARMLSAGYRVS 708
Qy 725 PDGKTGYVIVKGGGRISORAMPYFMVDPDSQIDVYVVDHQIIPAAIRILGYFGEITEKKL 784
Db 709 KGKIGYVIVKGGGRISORAMPYFMVDPDSQIDVYVVDHQIIPAAIRILGYFGEITEKKL 784
Qy 785 KASATGQKTLDFDLAK 800
Db 769 K---TGGVDILSFFKK 781

RESULT 4
ID Q8ZVU2 PRELIMINARY; PRT; 785 AA.
AC Q8ZVU2;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE DNA-directed DNA polymerase (B3).
GN PAE2109.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TM2 / ATCC 51768 / DSM 7523;
EX PubMed=11792869;
RA Fitz-Gibbon S.F., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
aerophilum";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AB009857; AAL63952.1; -.
DR InterPro; IPR006172; DNA pol B.
DR InterPro; IPR006134; DNA pol B dom.
DR InterPro; IPR006133; DNA pol B_exo.
DR InterPro; IPR004578; Pol2.
DR Pfam; PF00136; DNA pol B; 1.
DR Pfam; PF03104; DNA pol B_exo; 1.
DR PRINTS; PR00106; DNAPOLB.

```

```

DR SMART; SMO0486; POLB; 1.
DR TIGRFAMs; TIGR00592; pol2; 1.
DR PROSITE; PS00116; DNA POLYMERASE B; 1.
KW DNA-directed DNA polymerase; Complete proteome.
SQ SEQUENCE 785 AA; 89547 MW; 8395EA18C7804F1F CRC64;

Query Match      51.5%; Score 2156; DB 17; Length 785;
Best Local Similarity 53.7%; Pred. No. 7.2e-140;
Matches 428; Conservative 131; Mismatches 210; Indels 28; Gaps 9;

QY 9 LDSSVEYVKGSPQVINGIAENGERVVLIDRSFPYFALLAPGADPKQVQAQIRALSRP 69
DB 8 LDATSYVGGVPEVRFIFGISEGSDRVVVDVRRFPFYFA-DCPACDPESVRSQGRVA-P 65
QY 69 KSPIIGVDDDKKVFGRPRRLVRLTLPFAVREVLKVNVDGVLEADIFAFAYL 128
DB 66 VEEVAVB---RRVLGRPSRLKIVARPEVDVRLREAAALPGVSGYEAIDIRFYKYM 122
QY 129 IDHOLFPTTYRVEABLENKMGFRVDKVLVKSRLPELYGEA---LAPTKLPDLRLA 184
DB 123 LDMGVVPCSWNTVDAEATGKLG-----NLPVYKVAEMGGVTEGFPPLAVLA 170
QY 185 FDIENVYKQSGPRPERDVIIVIAKTDGDEVLFIAEGKDDKFIREFVYVKRYDPII 244
DB 171 FDIENVYRGTPDPLRDVILLAVQASDGRVFEVFEASGRDDRSVLSRFDIVREFDPVI 230
QY 245 VGYNNHFDWPLRRARILGILDVTRRVCAEPTTSVGHVSVPGRNLVDLYDYAEMP 304
DB 231 VGYNSQFDWYLAEARALGIPKVD-RUGARQQSVYGHWSVTGRANDVLINIVDEFP 289
QY 305 EIKTKSLEVAAYLGWKKSRVIVNWEIPDYWDPKRPLLLQYARDVRAFYGLAEK 364
DB 290 EIKLKTLDRAVEYFGVMKREBVLVPGHKIYEYVRDQGRPLQYVVDVVKSTYGLAEK 349
QY 365 ILPAIQLSYTGLDQVGAWSGFRLEWLIIRAFKMKELVNRVERPEETVYGAIVL 424
DB 350 LPLFIQLSYSGSLPLDQVAAASVGNRVEMMLLRYAYRLOGEVAPNREERYEPYKGAIVL 409
QY 425 EPLRGVHENIAVLFSSMYNIMIKYVNPDTLVRPGCKGCGCEWAPVGRFRRCPP 484
DB 410 EPRGLYSDVLALDFSSMYNIMIKYVNPDTLVRPGCKGCGCEWAPVGRFRRCPP 469
QY 485 GFYKTVLERLELRKRVRAEMKKYPPSPPEYELDLDERQKALKVLANASYGMWGSARWY 544
DB 470 GF1PLVLRQLTELRKRVREELKYPSPSPPEYELDLDERQKALKVLANASYGMWGSARWY 529
QY 545 CRECAKAVTAGRHILRTAINIARKLGLKVLGYGTDLSFYTYDPEKVENIKIKELGF 604
DB 530 KKEVAESTAPARAILDVIYARAGILVYIGTDSLFLVKSGD-VEKLKVVESYKGI 598
QY 605 EIKLEKVKYKRLFFTEAKRYAGLLEDGRIDIVGFEAVRGDWCELAKEVQTKVVEIVLKT 664
DB 589 DIKIDKDYTVLFTFAKRYAGLLRDGRIDIVGFEVVRGDSWSELAKEVQLRVIELIITSR 648
QY 665 EVN---KAVEYVKIKLEBQKVPTEKLVITKLSKRLSEYTERPHVVAKRMLSAG 720
DB 649 DVSEARQKVKYRGVIDKLNRYEVDLDDIITKTLDELKELDAKAYPPHVAAILLKRG 708
QY 721 YRVSPGKIGVIVKGGGRISORAMPYFMWKDPSQIDVTVYVDHQLIPALRLILUGGIT 780
DB 709 YKVGKGTIGVIVKGGKESERAVPVIFIDDIKIDLDYVVERQVIPAALRIAETVIGK 768
QY 781 EKLKASATGQKTLDF 797
DB 769 EGDLEK-TGRSERTLLDP 784

RESULT 5
ID Q9P9M2 PRELIMINARY; PRT; 785 AA.
AC Q9P9M2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

```

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Family B DNA polymerase (EC 2.7.7.7).
GN POLB3.
OS Pyrobaculum islandicum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=2277;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GEO 3;
RK MEDLINE=20100754; PubMed=10633098;
RA Kaehler M., Antranikian G.;
RT "Cloning and Characterization of a Family B DNA Polymerase from the
RT Hyperthermophilic Crenarchaeon Pyrobaculum islandicum.";
RL J. Bacteriol. 182:655-663 (2000).
CC -1- CATALYTIC ACTIVITY: 2' DEOXYNUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE
CC + {DNA} (N).
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
DR EMBL; AF195019; AAF27815.1; -.
DR HSSP; P56689; LTGO.
DR InterPro; IPR006172; DNA_pol_B.
DR InterPro; IPR006134; DNA_pol_B_dom.
DR InterPro; IPR004578; Pol2.
DR Pfam; PF00136; DNA_pol_B; 1.
DR Pfam; PF03104; DNA_pol_B_exo; 1.
DR PRINTS; PR00106; DNA_POLB.
DR SMART; SM00486; POLB; 1.
DR TIGRFAMs; TIGR00592; pol2; 1.
DR PROSITE; PS00116; DNA POLYMERASE B; 1.
KW DNA replication; DNA-binding; DNA-directed DNA polymerase;
KW Transference.
SQ SEQUENCE 785 AA; 89748 MW; 10FB8B66A8F3730D CRC64;

Query Match      50.8%; Score 2129.5; DB 1; Length 785;
Best Local Similarity 53.4%; Pred. No. 4.8e-138;
Matches 429; Conservative 141; Mismatches 195; Indels 39; Gaps 13;

QY 9 LDSSVEYVKGSPQVINGIAENGERVVLIDRSFRPVFY---ALLAPGADPKQVQAQIRAL 65
DB 8 LDITAVGSGVFEIRIFGILSSGERVVLIDRSFKFYFVDCVACBFA-----LKTAL 60
QY 66 SRPKSPIIGVEDDKKVFGRPRRLVRLTLPFAVREVLKVNVDGVLEADIFAFAM 125
DB 61 SR-VAPIDDDVQIVERRFLGRSKKELKVIKFIKEDVKUREAAMSIFRSGVYEAIRFYM 119
QY 126 RYLIHDHLPFTTYRVEABLENKMGFRVDKVV---VLVKSRLPELYG---EALAPTCLPDLR 181
DB 120 RYIMDMGVVPCSWNTVDAEVEE-----GRLGGITFYVWSQ-----WYIGDEGFPSP 166
QY 182 ILAFDIEVYSKQSGPRPERDPIVIAVKTDDGVLEFIAEGKDDKFIREFVYVKRYDP 241
DB 167 VNAFIEVYVNEGSDPDPIDPVVMLAIKINDGHEEVFEASGKDDGVVAFVDFIRSDP 226
QY 242 DIIVGNNHFDWPLRRARILGILDVTRRVGAETTSVGHVSVPGRNLVDLYDYAE 301
DB 227 DWIVGNSNGFDWPLVREAKAVGVPKVD-RLSNPPQQSVYGHWSIVGRANVDLYNIVE 285
QY 302 EMPEKIKSLEBAVYLGWKKSERVIVNWEIPDYWDPKRPLLLQYARDVRAFYGL 361
DB 286 EPEFKLKTLDRAVEYFGVMKREERVLIPGHKIYEYWKDPNKRPLLRKLYLDDVSTLGL 345
QY 362 ASKILPFAIQLSYTGLDQVGAWSGFRLEWLIIRAFKMKELVNRVERPEETVYGA 421
DB 346 ADKLFPFLIQLSSVGLPDDQVAAASVGNRVEMMLLRYAYRLOGEVAPNREERYEPYKGA 405
QY 422 IVLEPLRGVHENIAVLFSSMYNIMIKYVNPDTLVRPGCKGCGCEWAPVGRFRRCPP 481
DB 406 IVLEPKPGMYEDVLVDLDFSSMYNIMIKYVNPDTLVRPGCKGCGCEWAPVGRFRRCPP 465
QY 482 CPFGFKTVLERLELRKRVRAEMKKYPPSPPEYELDLDERQKALKVLANASYGMWGS 541
DB 466 SPFGFVPQVQLSLVELKAVREAEAKKYPDPSPFEKILDERQKALKVMAATVYGLWVGA 525

```

```

QY 542 RWCCECAKAVTAMGRHLIRTAINTARKLGLKVIYGDTSLSFVTVDPKPVNFIKKEE 601
DB 526 RMYKEVAVESVAFARAILKDVIOEARRLGIVVYGTDSLFVKHGD-VDKLKIYVEEK 584
QY 602 LGFEIKLKVYKRLFFTEAKKRYAGLLEDGRIDIVGFEAVRGDMCELAKEVQTKVVEIVL 661
DB 595 YGIDIKVDYAKVLFTFAKKRYAGLLEDGRIDIVGFEVVRGDMSELAKDQVLRVIEIIL 644
QY 662 KTSEYNKA-----VEVRKIVKLEKGVPIEKLVITWKLSKRLSEYTTTEAPHVVAAKML 717
DB 645 KSRDIVEARHGVIKIREIIRIELKNYKNIDDLIIWKTLDELDKELDYKAYPEHVHAAQILK 704
QY 718 SAGYVSPGDKIGYIVKGGGRISORAMPYFMVKDPSOIDVTVYVDHIOIIPAAIRILGYF 777
DB 705 RHGYVGGTIGYIVKGGKSERALPYILLDDIKIDIDYIERQIIPAAUKIAEVI 764
QY 778 GITEKLLKASATG--QKTLFDFIA 799
DB 765 GVKESDLK--TGRMERSLLDELS 785

RESULT 6
Q9P9K4 ID Q9P9K4 PRELIMINARY; PRT; 775 AA.
AC Q9P9K4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN DNA POL.
OS Pyrococcus glycovorans.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=74610;
RN [1]
RP SEQUENCE FROM N.A.
RA Quereillon J., Cambon M.A., Lesongeur F., Forterre P., Barbier G.;
RT "DNA polymerase genes organisation of species belonging to
Thermococcales and phylogenetic implications.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE
+ {DNA}(N).
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
DR EMBL; AJ004834; CAB81803.1; -.
DR HSP; F56689; ITGO.
DR InterPro; IPR006172; DNA_pol_B.
DR InterPro; IPR006134; DNA_pol_B_dom.
DR InterPro; IPR006133; DNA_pol_B_exo.
DR InterPro; IPR004578; pol2.
DR Pfam; PF00136; DNA_pol_B; 1.
DR Pfam; PF03104; DNA_pol_B_exo; 1.
DR PRINTS; PR00106; DNAPOLB.
DR SMART; SM00486; POLBC; 1.
DR TIGRFAMs; TIGR00592; pol2; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW DNA replication; DNA-binding; DNA-directed DNA polymerase;
KW Nucleotidyltransferase; Transferase.
SQ SEQUENCE 775 AA; 90492 MW; 677264920D770C0F CRC64;

Query Match 30.3%; Score 1268; DB 1; Length 775;
Best Local Similarity 36.5%; Pred. No. 1.1e-78;
Matches 304; Conservative 154; Mismatches 278; Indels 98; Gaps 23;

QY 8 VLDSSYVGVKPEQVVIINGIAENGERVVLIDRSFRFYFALAPADPKQV----AQIR 63
DB 2 ILDADYITDGGPIIRIFK-KENGFKVEYDENFRPIYALLKDDSQIDEVKVITAERHG 60
QY 64 ALSRPKSPIIGVEDDKRYKFGPRVRLIRTVLPEAVREYRELKVNVDGVEDVLEADIRF 123
DB 61 KIVR----IVDVEKVKKFLGRPIEVWKLIFYFHPQDQVPAIRDKIREHPAVVDIPEYDIPF 116

```

```

QY 124 AMRYLIDHLDPPFTWYRVAEAPLENKMGFRVDKVLVKSRPEPLYGEALAPTKLPDLIRL 183
DB 117 AKRYLIDKGLI-----FWEGD-----EELKLL 138
QY 184 AFDIEVSKSGSPRPDPVIVIAVKTDDGDEVL-----FIAEGKDDKPKIRFEVVEY 235
DB 139 AFDIETLYHEGE-BFAGKPIIMISYADEGAKVITWKVKDLPVYEVVSSREMIKFLKV 197
QY 236 VKRYDDPILIVGYNHNFDPYILLRRARILGILKIDVTRVGAETTSVGH-----VSPGRL 292
DB 198 IREKDPDVIITYNGDSFOLPYLVKRAEKLGIKPLGRD-GSEPKMORLGMTAVEIKRI 256
QY 293 NVLDYVAEEMPEIKISLEAEVYLGVMKKSRIINWWEIIPDYDDDPKRPILLQYAR 352
DB 257 HFDLYHVIRTIINLPYITLBAVYEAI-FGAPKEKVIAH--EIAEAMETGKLRVAKYSW 313
QY 353 DDVRAVYGLAEKILPFAIQLSYVTGLDQGVAMSVGFRLEWYLIIRAAFRKELVFNKVE 412
DB 314 EDKAVTVYELGREFPFMEQQLSRVGGPLMDVSSSTGNLVEWYLLRKAYERNELANPKPD 373
QY 413 RPE-----ETYRGAIVLEPLRGVHENIAVLDFSSMYPNIMIKYVGPDTLVRPGEKCGE 466
DB 374 EREYERRLRESYAGGYVKEPEKGLWGLVSLDFRSLYPSIIITHNVSPDTLNEE----- 427
QY 467 CGCWE--APEVKHRRPFCPPGFFKTVLRELRLKRVRAEMKKYPPDSPYRLLDERQK 523
DB 428 -GCWEYDVAPEVRHKFCDFPGFTPSLLKLLDERQEIKKRMKA-SKDPIEKKMLDYRQR 485
QY 524 ALKVLNANASYGVMGSGARWYCRECAKAVTAMGRHLIR-TAINIARKLGLKVIYGTDSL 582
DB 486 AIKILANSYGYGYKARWYCKEACSVTAMGREVIEFVRKLEEKFGFKVLYIDTDL 545
QY 583 FVTV---DPKVE---NFIKIKEELG--FEIKLKVYKRLFTTAKGYAGLLEDGRI 633
DB 546 YATIPGAKPEIEIKKALEFVEYINAKLPGLELEYGFYVRGFEV-TKKYALIDSEGI 604
QY 634 DIVGFEAVRGDMCELAKEVQTKVVEIVLKTSEYNKAVEYVRKIVKLEEKGVPIEKLV 693
DB 605 ITRGLEIVRDMSEIAKETQALEAILKHGNEVAVKIVKETEKLKSKYEIPEKLVY 664
QY 694 KTLKRLSEYTTTEAPHVAAKRLMSAGYRVSPGDKIGYIVVGGGRISORAMPYFMVK-- 751
DB 665 EQITRPLHEYKAIGPHVAVAKRLAAGVKVPGWIGYIVLRGDPISKEA---ILABEF 721
QY 752 DP--SQIDVTYVVDHIOIIPAAIRILGFGITEKLLKASATGKOTLFDPLAKKK 803
DB 722 DPRKHKYDABYYIENQVLPVAVRILEAFGYRKEDLRWQKTQTGLTAWLNVKKK 775

RESULT 7
Q9HH06 ID Q9HH06 PRELIMINARY; PRT; 775 AA.
AC Q9HH06;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN POL.
OS Pyrococcus glycovorans.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=74610;
RN [1]
RP SEQUENCE FROM N.A.
RA Quereillon J.J.E., Cambon M.A., Lesongeur F., Barbier G.;
RT "Thermococcales taxonomy and phylogeny based on the comparative use of
16S rDNA, 16S-23S rDNA intergenic spacer and family B DNA polymerase
genes.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE
+ {DNA}(N).
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.

```



```

QY 6 FTVLDSSEYVYKKE-POVILWIAENGERRVVLIDRSRPYFYALLAPGADPKQVAQRIR- 63
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 5 FOILDADYEVNDGSPVIRLFRGADGSKSVCCFVDFEPFY--LKASGDLHVARLIKD 62
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 64 ALSRPKSPILIGVEDDKRYKFORPRRVLIRITVLPAAVREYR-ELVK-----NVDGVEDVL 117
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 63 TFEQVKVVEIVEKEPEYQYKTKKEMLRVTRLPKDVPEIRDEILKIRDVLRAEGDQVY 122
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 118 EADIRFANRYLIDHDLFPFTYRVEABELE--NMGW-----FRVDKYVLVK--S 162
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 123 ESDILFRNRFLIDRALGGMVWVSAEGKVPDVRVILGAGSAWRSRCENFACDSAVLASGLK 182
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 163 RPEPLYGALAPTCLPDLIRILAFDIEVYKQS--PRPRDPIVIAVKT----- 211
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 183 RVENL--AIAP-----LKYLAFDIECLPLDGHPSPDVSIIISFSEFYKCHTKLI 234
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 212 -----DGEVLFTIAGKDDRRKPIREFYVYKRYDDPIIVGYNHNDPWLRLRAR 262
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 235 LLAKPAAGMDGD---VLSCMDYEMLNKFEIICEYDPIVAGYHQDQDPIYITERVK 290
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 263 ILGKILDVTRRVGAEPFTTSVGHVSP-----GRLNVDLYDAEEMPEIK 307
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 291 ALVAK-----GETINSVVRGDSPIGYKFKGLITRTKMKGRVVDVALPLVRRAPSLK 342
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 308 IKSLEEVAEYLGVMKKSERVIINWWEIPDYDDP--KKRPLLLQYARDVVRATYGLA--EK 364
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 343 QYTLRAVSKEL--LSREKLDVPPLEMEHNDGDKFRKFVDYARRDSELALSLVLELR 399
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 365 ILPAIQLSVYTGPLDQV---GAMSGVRLIEWLIRAAFOKMLNVRNVRPE----- 415
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 400 LLDKYIALAQVSGSLQLQIVDGGQTSM---VETLLLR--EFLGKO---RVILPKPGDEL 451
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 416 -ETV-----RGAIVLEPLRGVHENIAVLDFSSMYPMIMIKYVGPDTLV---RPGKRC- 464
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 452 AERDMSDLKGGEVLEPKGLENNVLIDLYKSYPTTMMAHNECYTTVVTRDPDQKTI 511
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 465 -----GEGCWEAPEVKEHRRCPFGFKTVLERLELRKRVRAEMKKYPPDPSPEYRLLD 519
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 512 KPSPGGE---FVPEV---FR---GIVPSILEDLNKRKGDTKKMKR--TSDENEHRLVD 560
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 520 ERQALKVLNASGYGWSGARWYCRECAKAVTAWGR--HLIRTAIARKLG----- 571
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 561 ATOLAIKILLNSFYGYGARARLYSLTLANAVTSFGRSNTLNTROLINGRIGVLRNS 620
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 572 -----LKVYGDTSLEFV-----TYDPEKV---ENFIKIKEEL 602
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 621 AALLIEBAGKLSPOQRIVELSVAYGDTDSVFVHCKAKGDLSELSVGNRLSEIVASL 680
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 603 --GFEIKLEKYKXBLFFTEAKRYA-GLLE-----DGRIDIVGEAVRGDWCCLAKEVQ 653
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 681 POPMELEPESVAKRALLI--AKRYALWLFEPNRSNGWENKIKVKGWETVRDWCCLTSITL 739
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 654 TKVVEITVLTSEVNKAVYVYKI---VKELEGKVP---IEKLVIWKTLSKRLSEYTTAP 708
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 740 NRVLFEVLIEGDVKAHEVRKAVSDVRNLDPGKDAGIIEKLVLTTLTRKADSYKKNQP 799
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 709 HWAAKRLMS--AGYRVSPGDKIGYVIVKGGGRISQRAW--PYFMVKDPSQIDVTYVDHQI 766
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 800 HLTVAENLKKTKTGMPSIGTRIPFVITAGKGLFVDRAEDPDYVRENNVPIDVDYVYKQI 859
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 767 IPAALRIILGYGITEKKLKASATGQKTLDFLAKK 801
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 860 LPPVERILLEVGVKMSLDFDAK--QKGLFDFEVKK 893
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

RESULT 12

```

ID Q96H98 PRELIMINARY; PRT; 1107 AA.
AC Q96H98
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Polymerase (DNA directed), delta 1, catalytic subunit (EC 2.7.7.7)

```

```

DE (125kd) (DNA polymerase).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=Lymph;
RA Strausberg A.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE
CC + {DNA} (N).
CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC DIFFERENT REACTIONS OF DNA SYNTHESIS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
DR EMBL; BC008800; AA08800.1; -.
DR InterPro; IPR0061172; DNA_pol_B.
DR InterPro; IPR0061134; DNA_pol_B_dom.
DR InterPro; IPR0061133; DNA_pol_B_exo.
DR InterPro; IPR004578; Pol2.
DR Pfam; PF001136; DNA_pol_B; 1.
DR PRINTS; PF03104; DNA_pol_B_exo; 1.
DR SMART; SM00486; POLBc; 1.
DR TIGRFS; TIGR00592; pol2; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B_1.
KW DNA replication; DNA-binding; DNA-directed DNA polymerase;
KW Transference.
SQ
Sequence 1107 AA; 123611 MW; 10A8A3271916EDBA CRC64;
Query Match 18.2%; Score 760.5; DB 4; Length 1107;
Best Local Similarity 28.7%; Pred. No. 1.8e-43;
Matches 255; Conservative 144; Mismatches 320; Indels 169; Gaps 33;
QY 17 GKPEQVIVWIAENGERRVVLIDRSRPYFYALLAPGADPKQVAQRIR-----ALSREP--- 69
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 120 GSVPLRAFGVTDGFSVCHIHGFPAYFYTPAPGFGFEGHMGDLQRLNLAISDRSG 179
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 70 -----SPILGVEDDKK-----YGR-PRRVLRITVLPVAVREYRLKXNDGVEDV--- 116
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 180 RELTGPAVLAVELCSRESMFGYGHGSPFLRTALPRLVAPARLLGQIRVAGLTP 239
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 117 ----LRADIRFANRYLIDHDLPEFTVYRVEAELENKMGFRVDKVL-----VKSRP 164
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 240 SPAPYEAIVDFEIRFMVDTDIVGCNMLELPAGKYALRLKEKATQCLEADVMSDVVSHP 299
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 165 EPLYGALAPTCLPDLIRILAFDIEVYKQS--SPRPERDFVIVI---AVKTDGDEVLFIA 220
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 300 PEGPWQRIAP-----LRVLSFDIECAGRKGIPEPERDPVIQICSLGLAWGEPEPFLRLA 354
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 221 -----EGKDDRKPIREFYVYKRYDDPIIVGYNHNDPWLRLRA--- 261
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 355 LTLRCPAIIAGKAVQSYKEEDL--LQAWSTFIRIMDPDVIITQNFDPYLSRAQT 412
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 262 -----RIIGIKLDV-----TRRVGABETTSHVGHVSVGRNLVDLYDAEEMPE 305
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 413 LKVTQTFPFLORVAGLCSNIRDSFSQKQTRDKV-----VSMVGVRQMDMLQVL--LRE 466
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 306 IKIKS--LBEVA--EYLGVNKKSERVIINWWEIPDYWDPPKCRPLLIQYARDVVR---ATY 359
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 467 YKLRSYTLNANVSPHFLGKEQEDVQHSI-----ITD-----LONGNDQTRRLAY 511
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 360 GLAEKILPP-----ATQLSVYTGDLDDQVAMSVGRFLRWYLRFAAKMKELVP 408
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 512 CLKDAYLPLRLRLERLMVLNVAEMARVTVPLSYLLSRGQVKKVSQLRQAMHEGLMP 571
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 409 NRVERPEETVYRGAVILEPLRGVHE--NIAVLDFSSMYPMIMIKYVGPDTLVPRG--EKCG 465
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 572 VKSEGGEDYTGATVIEPLRGYVDVPIATLDFSSLYPSIMMAHNLCTYTLRPGTAQKLG 631
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 466 ECGCWEAPEVKEHRRCPFG--FFKT-----VLERLLELRKRVRAEMKKYPPDPSPE 514
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```


Db	632	-----LTDQFIRPTGDEFVKTSVRKGLLPQILENLSARKRAELAK-ETDPLR	682
Qy	515	YRLDERQKALKVLANSYGYMGWSGARWYCRCAKAVTANGHLLI-----RTAIN	565
Db	683	RQVLDRQLALKVANSVYFTGAQVGLPCLEISQSVTGFGRQMIETKQLVESKYTYE	742
Qy	566	IARKGLKVIYGTDSLSFTYDEKVENFIKKEELGF-----EIKLEKVKRLF	616
Db	743	NGYSTAKVYGTDSVNCRFVSSVAEAMALGREAADWVGHPPSPIRLEFEKVFP-Y	801
Qy	617	FTEAKKYAGLL-----EDGRIDVGFVAVRGMCELAKEVQTKVVEIVLTKTSEYNKAV	670
Db	802	LLISKYVAGLLSSRPDAHDMCKGLEAVRDNCPVLANVYASRLRLIDRDEGAV	861
Qy	671	EYVRKIVKELEEGKVPKEIKLVIMKLSKRLSEYVTEAPHVAAKRM--LSAGVYVSPGDK	728
Db	862	AHAQDVISDLNCRIDISQVITKELTRAAADYAGKQAHVELAERMRKDPGAPSGLDR	921
Qy	729	IGYVIV--KGGGRISQAMPYFVMDPSQIDVTYVYDHIIPAAALRI	773
Db	922	VPYVILSAAGVAAMKSEDFLVIEHSLPIDTQYILEQQLAKPLRLI	969
RESULT 13			
Qy	Q8C2N0	PRELIMINARY; PRT; 1107 AA.	
AC	Q8C2N0		
DT	01-OCT-2002	(Tremblrel. 22, Created)	
DT	01-OCT-2002	(Tremblrel. 22, Last sequence update)	
DT	01-MAR-2003	(Tremblrel. 23, Last annotation update)	
DE	Polymerase (DNA directed), delta 1, catalytic subunit (EC 2.7.7.7)		
DE	(DNA polymerase).		
GN	FOLD1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Rieder M.J., Livingston R.J., Braun A.C., Montoya M.A., Chung M.-W.,		
RA	Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,		
RA	Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;		
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.		
CC	-!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE		
CC	+ [DNA] (N).		
CC	-!- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:		
CC	ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR		
CC	DIFFERENT REACTIONS OF DNA SYNTHESIS (BY SIMILARITY).		
CC	-!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.		
DR	EMBL; AY129569; AAM76971.1; -		
DR	InterPro; IPR006172; DNA_pol.B.		
DR	InterPro; IPR006134; DNA_pol.B_dom.		
DR	InterPro; IPR004578; Pol2.		
DR	Pfam; PF00136; DNA_pol.B.1.		
DR	PRINTS; PR00106; DNAPOLB.		
DR	SMART; SM00486; POLBC.1.		
DR	TIGRfam; TIGR00592; Pol2; 1.		
DR	PROSITE; PS00116; DNA_POLYMERASE_B.1.		
KW	DNA replication; DNA-binding; DNA-directed DNA polymerase;		
KW	Transferase.		
SQ	SEQUENCE 1107 AA; 123630 MW; 9D04D34AB4AE810 CRC64;		
Query Match			
Best Local Similarity 18.2%; Score 760.5; DB 4; Length 1107;			
Matches 255; Conservative 144; Mismatches 320; Indels 169; Gaps 33;			
Qy	17	GKEPQVILGIRANGBRVILDRSPFRPYVALLAPGADPKQVAQIR-----ALSRPX---	69
Db	120	GSVFLRAGVIVDEGSVCCHTGFAPYFTTAPFGFHEHMGDLQRLNLAISDRSGG	179
Qy	70	-----SPITIGVEDDKRK-----YFGR-PRVLRIRTVLPEAVREYRELKVNQVGDV---	116
Db	180	RELTGPAVLAVELCSRESNFGYHGHPSPFLRITVALPRLVAPARLLAQGIRVAGLGT	239

Qy	117	-----LEADIRAMRYLIDHDLFPPTWYRVEAEPLNKMGRVVDKVL-----VKSRRP	164
Db	240	SPAPYANVDFEIRPWDTDIVGCNWLPLPAGKALRLKEKATCCQLEADVLSVDVSH	299
Qy	165	EPLYGEALAPTKLPDLRLAFDIEVYSKQ--SPRPDPVVI---AVKTDGDEVLPIA	220
Db	300	PEGPWQRIAP-----LRVLSFDIECAGRKGIFFPERDPVIOICSLGLRWGEPEPFLKA	354
Qy	221	-----EGKODRKIRFVFEVVKRYDPIIIVGYNHNFDPYLLRLA--	261
Db	355	LTLRCPAPILGAKVQSYKEEDL--LQAWSTFIRIMDPDVTGNYIQNFDPYLLISRAOT	412
Qy	262	-----RILGKLDV-----TRRVGAEPSTSVGHVSVPGRANLDVLYAEEMPE	305
Db	413	LKVQTFPFLGRVAGLCSNIRDSFSQKQTRDRTKV-----VSMGVRQVMDMLQVL--LRE	466
Qy	306	IKKS--LBEVA-EYLGVMKKSERVIINWMEIPDYWDPPKPKRLLQVARDVR---ATY	359
Db	467	YKLSVYTLNAVSHFLGEGEKEDVQHSI-----ITD-----LONGNDQTRRLAVY	511
Qy	360	GLAEKILPP-----AIQLSVYVTCGLPDQVGMVSGPRLSWILIRAAFKMKELVP	408
Db	512	CLKDAYLPLRLERLMVLNVAEMARVTGVLISLRSQQVVKVVSQQLRQAMHGLLMP	571
Qy	409	NRVERPEETIRGAIVLEPLRGVHE--NIAVLDFSSMYPMIKYNYGPDTLVRPG--BKQ	465
Db	572	VVKEGEGDYTGATVIEPLKGYDYDPIATLDFSSVPSIMMAHNLCTYTLRPPGTAQKLG	631
Qy	466	EGCWEAPEVVKHPRRCPPG--FFKT-----VLERLLSLRKVRBAEMKYPDPSP	514
Db	632	-----LTDQFIRPTGDEFVKTSVRKGLLPQILENLSARKRAELAK-ETDPLR	682
Qy	515	YRLDERQKALKVLANSYGYMGWSGARWYCRCAKAVTANGHLLI-----RTAIN	565
Db	683	RQVLDRQLALKVANSVYFTGAQVGLPCLEISQSVTGFGRQMIETKQLVESKYTYE	742
Qy	566	IARKGLKVIYGTDSLSFTYDEKVENFIKKEELGF-----EIKLEKVKRLF	616
Db	743	NGYSTAKVYGTDSVNCRFVSSVAEAMALGREAADWVGHPPSPIRLEFEKVFP-Y	801
Qy	617	FTEAKKYAGLL-----EDGRIDVGFVAVRGMCELAKEVQTKVVEIVLTKTSEYNKAV	670
Db	802	LLISKYVAGLLSSRPDAHDMCKGLEAVRDNCPVLANVYASRLRLIDRDEGAV	861
Qy	671	EYVRKIVKELEEGKVPKEIKLVIMKLSKRLSEYVTEAPHVAAKRM--LSAGVYVSPGDK	728
Db	862	AHAQDVISDLNCRIDISQVITKELTRAAADYAGKQAHVELAERMRKDPGAPSGLDR	921
Qy	729	IGYVIV--KGGGRISQAMPYFVMDPSQIDVTYVYDHIIPAAALRI	773
Db	922	VPYVILSAAGVAAMKSEDFLVIEHSLPIDTQYILEQQLAKPLRLI	969
RESULT 14			
Qy	Q8C2N0	PRELIMINARY; PRT; 1105 AA.	
AC	Q8C2N0		
DT	01-MAR-2003	(Tremblrel. 23, Created)	
DT	01-MAR-2003	(Tremblrel. 23, Last sequence update)	
DT	01-MAR-2003	(Tremblrel. 23, Last annotation update)	
DE	DNA polymerase delta 1.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NOD; TISSUE=Thymus;		
EX	MEDLINE=22354683; PubMed=12466851;		
RA	The FANTOM Consortium,		
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;		
RT	*Analysis of the mouse transcriptome based on functional annotation of		

```

RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK088310; BAC40275.1; -.
SQ SEQUENCE 1105 AA; 123776 MW; 2032377DD4722B1 CRC64;

Query Match 17.8%; Score 743.5; DB 11; Length 1105;
Best Local Similarity 27.9%; Pred. No. 2.6e-42;
Matches 246; Conservative 153; Mismatches 319; Indels 165; Gaps 33;

QY 20 PQVIWGIABNGERVVLDLRSFPPYFALLAGADPKQVAQIR-----ALSRLPK----- 69
Db 121 PILRAGVDTDEGSVCCHIQGFAPYFYPAPFGFAGHLSLQELNAAISRDQGGKEL 180
QY 70 --SPIIGVEDDKK-----YFGR-PRRLVIRTVLPEAVREYRELK---NVQVED----- 115
Db 181 SGPAVLATIELCSRESMFGYHGHGSPFLRITLALPRLMAPARRLELQGVRFGLTFPSFA 240
QY 116 VLEADIRFAMRYLIDHDLFFFTW-----YRVEAPLENKMGFRVDKYL-VKSRPEPL 167
Db 241 PYEANDVFIRFMDVADIVGCNWLLELPAGKYVRAEKKATLCQLEVDVLSWDVISHPPRG 300
QY 168 YGEALAPTCLDLRIILAFDIEVYSKQG-SPRPBRDPVIVI---AVKTDGDEVLFIA--- 220
Db 301 QWQRIAP-----LRVLSFDIECAGRKGIFFEPERDPVIOICSLGLWGEPPFLRLALT 355
QY 221 -----EGKDRKPIREVEVVKYDDPDIIVGNNHDPWYLLRRA----- 261
Db 356 RPCAPILGAKVQSYEREEDL--LQAWADFILAMDPOVITGYNIQNFEDLYLSRAQALKV 413
QY 262 -----RIIGIKLDV-----TRRYGABFTTSVHGHVSVPGLNVDLYDAEEMPEIKI 308
Db 414 DRPPFLGRVTVGLSNIIRDSFSQSRQVGRDSKV-----ISMVGRVQMDMLQVLLREHKLS 469
QY 309 KSLLEVA-EYLGVMKKSERVINWWDIPYDDPKRPLLLQYARDVR---ATYGLAEK 364
Db 470 YTLNAVSHFLGSEKEDVQHSI-----ITD-----LQNGNEQTRRLAVYCLKDA 514
QY 365 ILPPA-----IQLSVTVGLDQVGAMSVGRLEWVILIRAAFKMKELVPRVER 413
Db 515 FLPLRLERLMLVNNVEMARVTVGLYLLTRGQGVKVVQQLLRQAMQGLLMPVVKTE 574
QY 414 PEETVRGAIVLEPLRGVHE-NIAVLDFSSMYNIMIKYVNDPLVRPG--EKCGEGCW 470

AC Q91VT0;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-WAR-2003 (TRENBLrel. 23, Last annotation update)
DE DNA polymerase delta 1, catalytic domain (EC 2.7.7.7).
GN POLD1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -|- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE
CC + [DNA] (N).
CC -|- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC DIFFERENT REACTIONS OF DNA SYNTHESIS (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.
DR EMBL; BC009128; AA039128.1; -.
DR MGD; MGI:97741; PolD1.
DR InterPro; IPR006172; DNA_pol_B.
DR InterPro; IPR006134; DNA_pol_B_dom.
DR Efam; PF00136; DNA_pol_B_1.
DR PRINTS; PR00106; DNAPOLB.
DR SMART; SM00486; POLB; 1.
DR TIGRFAM; TIGR00592; pol2; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KM DNA replication; DNA-binding; DNA-directed DNA polymerase;
KW Transferase.
SQ SEQUENCE 1105 AA; 123744 MW; 20323690DD472406 CRC64;

Query Match 17.8%; Score 743.5; DB 11; Length 1105;
Best Local Similarity 27.9%; Pred. No. 2.6e-42;
Matches 246; Conservative 153; Mismatches 319; Indels 165; Gaps 33;

QY 20 PQVIWGIABNGERVVLDLRSFPPYFALLAGADPKQVAQIR-----ALSRLPK----- 69
Db 121 PILRAGVDTDEGSVCCHIQGFAPYFYPAPFGFAGHLSLQELNAAISRDQGGKEL 180
QY 70 --SPIIGVEDDKK-----YFGR-PRRLVIRTVLPEAVREYRELK---NVQVED----- 115
Db 181 SGPAVLATIELCSRESMFGYHGHGSPFLRITLALPRLMAPARRLELQGVRFGLTFPSFA 240
QY 116 VLEADIRFAMRYLIDHDLFFFTW-----YRVEAPLENKMGFRVDKYL-VKSRPEPL 167
Db 241 PYEANDVFIRFMDVADIVGCNWLLELPAGKYVRAEKKATLCQLEVDVLSWDVISHPPRG 300
QY 168 YGEALAPTCLDLRIILAFDIEVYSKQG-SPRPBRDPVIVI---AVKTDGDEVLFIA--- 220
Db 301 QWQRIAP-----LRVLSFDIECAGRKGIFFEPERDPVIOICSLGLWGEPPFLRLALT 355
QY 221 -----EGKDRKPIREVEVVKYDDPDIIVGNNHDPWYLLRRA----- 261
Db 356 RPCAPILGAKVQSYEREEDL--LQAWADFILAMDPOVITGYNIQNFEDLYLSRAQALKV 413
QY 262 -----RIIGIKLDV-----TRRYGABFTTSVHGHVSVPGLNVDLYDAEEMPEIKI 308
Db 414 DRPPFLGRVTVGLSNIIRDSFSQSRQVGRDSKV-----ISMVGRVQMDMLQVLLREHKLS 469
QY 309 KSLLEVA-EYLGVMKKSERVINWWDIPYDDPKRPLLLQYARDVR---ATYGLAEK 364
Db 470 YTLNAVSHFLGSEKEDVQHSI-----ITD-----LQNGNEQTRRLAVYCLKDA 514
QY 365 ILPPA-----IQLSVTVGLDQVGAMSVGRLEWVILIRAAFKMKELVPRVER 413
Db 515 FLPLRLERLMLVNNVEMARVTVGLYLLTRGQGVKVVQQLLRQAMQGLLMPVVKTE 574
QY 414 PEETVRGAIVLEPLRGVHE-NIAVLDFSSMYNIMIKYVNDPLVRPG--EKCGEGCW 470

RESULT 15
Q91VT0
ID Q91VT0 PRELIMINARY; PRT; 1105 AA.

```

```

Db 575 GGEDYTGATVIEPLKGYDVPATLDFSSLYPSIMMAHNLCTVTLLRPGAQKLG----- 629
QY 471 EAPE-----VKHRFRCPGPFKTVLERLLELRKVRRAEMKKYPPDPSPYRLLD 519
Db 630 LKPDEFIKTPTGDEFVKSVRK--GLLPQILENLLSARKRAKAEALQ-ETDPLRRQVLD 685
QY 520 EROKALKVLANASYGVMGWSGARWYCRECAKAVTAWGRHLJRTAINIARK-----L 570
Db 686 GRQLALKVSANSVYGTGAQVGLKPLCLEISQSVTGFGRQMIEKTKQLVESKYTVENGDA 745
QY 571 GLKVIYGTDSLFTYDPEKVENFIKIKEBLGP-----EIKLEKVVYKRLFFTEAK 621
Db 746 NAKVYVGTDSVMCRFGVSSVAEAMSLGREAAANWVSSHPPSPIRLEPEKVVFP-YLLISK 804
QY 622 KEYAGLLEDGR-----IDIVGFPAVRGDWCELAKVEQTKVVEIVLKTSEVKNKAVEYVRK 675
Db 805 KRYAGLLFSSRSDAHDKMDCKGLEAVRRDNCPLVANLVTSLSRLRILVDRDPDGAVAHAKD 864
QY 676 IVKELEEGKVPTEKLVIMKTLKRLSEYTTAPHVVAARM--LSAGYRVSPGDKIGYVI 733
Db 865 VISDLNCRIDISQLVITKELTRAADYAGQAHVELAERMKRDPGSAPSLGRVPYVI 924
QY 734 V---KGGGRISORAWPYFMVKDPSQIDVTYYVDHQIIPAALEI 773
Db 925 IGAAGVAAYMKSEDFLVLEHSLPDTQYVLEQQLAKPLLRI 967

```

Search completed: November 25, 2003, 14:59:52
Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 14:52:23 ; Search time 48 Seconds
(without alignments)
2655.363 Million cell updates/sec

Title: US-10-034-849-2
Perfect score: 4188
Sequence: 1 MTEVFTVLDSSYEVVGKEP.....LKASATGQKTLFDPLAKSK 803

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 19Jun03:*

- 1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	4188	100.0	803	20	AAV00936
2	4188	100.0	803	23	AAE22118
3	4188	100.0	803	24	ABG73155
4	4188	100.0	803	24	ABG73160
5	4188	100.0	803	24	ABG73162
6	4164	99.4	803	23	AAE22119
7	4164	99.4	803	24	ABG73161
8	4164	99.4	803	24	ABG73163
9	4122	98.4	799	23	AAE22120
					P. fumarius DNA po
					Pyrolobus fumaria
					Pyrolobus fumarius
					Desulphurococcus t
					Pyrolobus thernost
					Pyrolobus fumaria
					Aquifex pyrophilus
					Pyrolobus thernost
					Pyrolobus fumaria

Pyrodictium occult
P. occultum DNA po
P. abyssi DNA poly
Pyrobaculum island
A. fulgidus AF0497
A. fulgidus AF0497
A. lithotrophus
Archaeoglobus lith
Archaeoglobus therm
P. kodakaraensis K
Heat-resistant Pfu
P. kodakaraensis K
P. kodakaraensis K
Pyrococcus kodakar
Pyrococcus kodakar
P. kodakaraensis K
Mutant KOD DNA pol
Heat-resistant Pfu
Heat-resistant Pfu
Amino acid sequenc
T. profundus therm
Heat-resistant DNA
Wild type Pyrococc
Pyrococcus furiosu
Wild-type P. furio
P. furiosus (Pfu)
P. furiosus DNA po
Pyrococcus furiosu
Thermococcus gorgo
Mature DNA polymer
KOD1 thermostable
Thermococcus pepto
Mutant KOD DNA pol
Mutant KOD DNA pol

ALIGNMENTS

RESULT 1
AAV00936
ID AAV00936 standard; Protein; 803 AA.
XX AC AAV00936;
XX DT 02-JUN-1999 (first entry)
XX DE P. fumarius DNA polymerase 1PY2 protein sequence.
XX DE DNA polymerase; thermophilic bacteria; DNA synthesis.
XX KW Pyrolobus fumarius.
XX OS WO9907837-A1.
XX PN 18-FEB-1999.
XX PD 06-AUG-1998; 98WO-US17152.
XX PF 06-AUG-1997; 97US-0907166.
XX PR (DIVE-) DIVERSA INC.
XX PA Callen W, Mathur EJ;
XX PI WPI; 1999-180490/15.
XX DR N-PSDB; AAX27283.
XX DR DNA polymerases from extremely thermophilic bacteria - useful for
XX PT DNA synthesis
XX XX

PS	Claim 1; Fig 2; 72pp; English.	
XX	This sequence is a DNA polymerase of the invention, that was isolated from a thermophilic bacteria. The polymerases are used in DNA synthesis and as immunogens to raise antibodies (useful for affinity purification and to screen for related enzymes). Fragments of the DNA encoding the polymerases are used as probes to isolate related or full-length sequences and to produce the recombinant polymerases. The polymerases catalyse DNA synthesis by the addition of deoxynucleotides to the 3' end of a polynucleotide chain, using a complementary polynucleotide strand as a template. The polymerases have optimum activity at over 60 deg. C and can renature and regain activity after exposure to temperatures above 70 deg. C.	
XX	Sequence 803 AA;	
SQ	Query Match 100.0%; Score 4188; DB 20; Length 803; Best Local Similarity 100.0%; Pred. No. 0; Matches 803; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MTEVFTVLDSSYEVGKPEQVIWIAENGGERVVLIDRSRPFYFALLAPGADPKQVAQ 60	
DB	1 MTEVFTVLDSSYEVGKPEQVIWIAENGGERVVLIDRSRPFYFALLAPGADPKQVAQ 60	
QY	61 RIRALSRPKSPIIGVEDDKRKYFGRPRVLRIRTVLPEAVREYRELKVNVDGVEDVLEAD 120	
DB	61 RIRALSRPKSPIIGVEDDKRKYFGRPRVLRIRTVLPEAVREYRELKVNVDGVEDVLEAD 120	
QY	121 IRFAMRYLIDHDLFPFTWYRVEAEPLNKGFRVDKVLVKSRRPEPLYGEALATPKLPDL 180	
DB	121 IRFAMRYLIDHDLFPFTWYRVEAEPLNKGFRVDKVLVKSRRPEPLYGEALATPKLPDL 180	
QY	181 RILAFDIEVYSKQSPRPERDPVIVIAVKTDDGDEVLFIAEGKDKRPIREFVEYVKRYD 240	
DB	181 RILAFDIEVYSKQSPRPERDPVIVIAVKTDDGDEVLFIAEGKDKRPIREFVEYVKRYD 240	
QY	241 PDIIYGVNNHFDWPLLRRARILGILKLDVTRVCAEPTTSVGHVSVPGLNVDLYDYA 300	
DB	241 PDIIYGVNNHFDWPLLRRARILGILKLDVTRVCAEPTTSVGHVSVPGLNVDLYDYA 300	
QY	301 EEMPEIKIKLSLEAEAYLVGMKKSERVIINWMEIPDYWDPPKRPPLLQYARDVRYTG 360	
DB	301 EEMPEIKIKLSLEAEAYLVGMKKSERVIINWMEIPDYWDPPKRPPLLQYARDVRYTG 360	
QY	361 LAEKILPFAIQLSYVTGLDQVGAMSVGRLEWYLIRAAFKMKSLVFNVRPERPETRYG 420	
DB	361 LAEKILPFAIQLSYVTGLDQVGAMSVGRLEWYLIRAAFKMKSLVFNVRPERPETRYG 420	
QY	421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYVNGPDTLVRPGKCGCGWEAPEVKKRFR 480	
DB	421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYVNGPDTLVRPGKCGCGWEAPEVKKRFR 480	
QY	481 RCPFGFFKTVLELLELRKRVRAEMKKYPPDSPYEVRLDDEKQALKVLNANASYGMWSG 540	
DB	481 RCPFGFFKTVLELLELRKRVRAEMKKYPPDSPYEVRLDDEKQALKVLNANASYGMWSG 540	
QY	541 ARMYCEKAVTANGRHILRTAINTARKLGLKVIYGDTSLSFVTVDPKVENFIKIKE 600	
DB	541 ARMYCEKAVTANGRHILRTAINTARKLGLKVIYGDTSLSFVTVDPKVENFIKIKE 600	
QY	601 ELGFEIKLEKVKYKRLPFTTEAKRYAGLLEDGRIDIVGFEAVRGDWCCLAKEVQTKVEIV 660	
DB	601 ELGFEIKLEKVKYKRLPFTTEAKRYAGLLEDGRIDIVGFEAVRGDWCCLAKEVQTKVEIV 660	
QY	661 LKTSVNVKAVEYVYKIVKELEGKVPFIEKLVINKTLSKRLSEYTTTAPHVAAKRMLSAG 720	
DB	661 LKTSVNVKAVEYVYKIVKELEGKVPFIEKLVINKTLSKRLSEYTTTAPHVAAKRMLSAG 720	
QY	721 YRVSFGDKIGYIVVKGGRISQRAWPYFMVKDPSQIDVTYVDHQIIPAAIRLTLGYFGIT 780	
DB	721 YRVSFGDKIGYIVVKGGRISQRAWPYFMVKDPSQIDVTYVDHQIIPAAIRLTLGYFGIT 780	
QY	781 EKKLKASATGQKTLFDLFLAKSK 803	

DB	781 EKKLKASATGQKTLFDLFLAKSK 803	
RESULT 2		
ID	AAE22118 standard; Protein; 803 AA.	
XX	AAE22118;	
AC	AAE22118;	
XX	25-JUL-2002 (first entry)	
DT	Pyrolobus fumaria DNA polymerase, 1PY2.	
XX	DNA polymerase; thermostable; enzyme.	
OS	Pyrolobus fumaria.	
XX	WO200220735-A2.	
PN	14-MAR-2002.	
PD	06-SEP-2001; 2001WO-US28007.	
XX	06-SEP-2000; 2000US-0656309.	
XX	(DIVE-) DIVERSA CORP.	
XX	Callen W, Mathur BJ, Short JM;	
PI	WPI; 2002-362247/39.	
XX	N-PSDB; AAD35187.	
DR	New thermostable polymerase useful for sequencing DNA, amplifying double stranded DNA, or incorporating a non-natural nucleotide or a nucleotide analog into a DNA molecule	
PT	Claim 42; Fig 1A-E; 161pp; English.	
XX	The invention relates to thermostable DNA polymerases having high temperature polymerase activity, such as those derived from Pyrolobus fumaria and nucleic acid molecules encoding such polymerases. Polymerases are useful for catalysing the formation or repair of a nucleic acid sequence and for modifying small molecules. They are also useful for sequencing DNA molecules, for preparing cDNA from mRNA, for amplifying double stranded DNA molecules and for incorporating non-natural nucleotides or nucleotide analogues into a DNA molecule. The present sequence is Pyrolobus fumaria DNA polymerase.	
CC	Sequence 803 AA;	
SQ	Query Match 100.0%; Score 4188; DB 23; Length 803; Best Local Similarity 100.0%; Pred. No. 0; Matches 803; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MTEVFTVLDSSYEVGKPEQVIWIAENGGERVVLIDRSRPFYFALLAPGADPKQVAQ 60	
DB	1 MTEVFTVLDSSYEVGKPEQVIWIAENGGERVVLIDRSRPFYFALLAPGADPKQVAQ 60	
QY	61 RIRALSRPKSPIIGVEDDKRKYFGRPRVLRIRTVLPEAVREYRELKVNVDGVEDVLEAD 120	
DB	61 RIRALSRPKSPIIGVEDDKRKYFGRPRVLRIRTVLPEAVREYRELKVNVDGVEDVLEAD 120	
QY	121 IRFAMRYLIDHDLFPFTWYRVEAEPLNKGFRVDKVLVKSRRPEPLYGEALATPKLPDL 180	
DB	121 IRFAMRYLIDHDLFPFTWYRVEAEPLNKGFRVDKVLVKSRRPEPLYGEALATPKLPDL 180	
QY	181 RILAFDIEVYSKQSPRPERDPVIVIAVKTDDGDEVLFIAEGKDKRPIREFVEYVKRYD 240	
DB	181 RILAFDIEVYSKQSPRPERDPVIVIAVKTDDGDEVLFIAEGKDKRPIREFVEYVKRYD 240	
QY	241 PDIIYGVNNHFDWPLLRRARILGILKLDVTRVCAEPTTSVGHVSVPGLNVDLYDYA 300	

Db 241 PDIIYGNVNHDPWYLLRRARILGKLDVTRRVGAETTSVGHVSVGRLNVDLYDA 300
 QY 301 EEMPEIKISLEVAEYLGVMKKSRRVIVNWEIPDYWDPPKRPPLLIQYARDVRYTG 360
 Db 301 EEMPEIKISLEVAEYLGVMKKSRRVIVNWEIPDYWDPPKRPPLLIQYARDVRYTG 360
 QY 361 LAEKILPRAIQLSYVTGLPDQVGAMSGFRLEWYLIRAAFKKELVNRVERPEETVRG 420
 Db 361 LAEKILPRAIQLSYVTGLPDQVGAMSGFRLEWYLIRAAFKKELVNRVERPEETVRG 420
 QY 421 AIVLEPLRGVHENIAVLFSSMYPNIMIKYVNPDTLVRPGEKCGECGCEWAPVKHRR 480
 Db 421 AIVLEPLRGVHENIAVLFSSMYPNIMIKYVNPDTLVRPGEKCGECGCEWAPVKHRR 480
 QY 481 RCPDGFKTVLRLLELRKRVRAEMKKYPPDPSPEYLLDEROKALKVLANASYGMWSG 540
 Db 481 RCPDGFKTVLRLLELRKRVRAEMKKYPPDPSPEYLLDEROKALKVLANASYGMWSG 540
 QY 541 ARWYCECAKAVTAMGRHLIRTAINTARKLGKLVYGDTSLVFYDPEKVENFIKIKE 600
 Db 541 ARWYCECAKAVTAMGRHLIRTAINTARKLGKLVYGDTSLVFYDPEKVENFIKIKE 600
 QY 601 ELGFEIKLEKVKRFFFTFAKKRYAGLLEDGRIDIVGEAVRGDWCELAKEVQTKVIV 660
 Db 601 ELGFEIKLEKVKRFFFTFAKKRYAGLLEDGRIDIVGEAVRGDWCELAKEVQTKVIV 660
 QY 661 LKTSVKNKAVEYVVKIVKLEBQKPIEKLIVTKLSKLEBYTTEAPHVAAKRWLSAG 720
 Db 661 LKTSVKNKAVEYVVKIVKLEBQKPIEKLIVTKLSKLEBYTTEAPHVAAKRWLSAG 720
 QY 721 YRUSPGKIGYIVKGGGRISQRAWFYVWVKDPSQIDVTYVVDHQIIPAAIRILGYFGIT 780
 Db 721 YRUSPGKIGYIVKGGGRISQRAWFYVWVKDPSQIDVTYVVDHQIIPAAIRILGYFGIT 780
 QY 781 EKKLKASATQKTLDFDLAKSK 803
 Db 781 EKKLKASATQKTLDFDLAKSK 803

RESULT 3
 ABG73155
 ID ABG73155 standard; Protein; 803 AA.
 AC ABG73155;
 XX
 DT 08-APR-2003 (first entry)
 XX
 DE Pyrolobus fumarius thermostable DNA polymerase protein.
 XX
 KW Enzyme; thermostable; DNA polymerase; DNA repair;
 KW polymorphism identification.
 XX
 OS Pyrolobus fumarius.
 XX
 PN US2002132243-A1.
 XX
 PD 19-SEP-2002.
 XX
 PF 06-SEP-2001; 2001US-0948369.
 XX
 PR 06-AUG-1997; 97US-0907166.
 PR 07-SEP-1999; 99US-0391340.
 PR 06-SEP-2000; 2000US-0656309.
 XX
 PA (CALL/) CALLEN W.
 PA (MATH/) MATHUR E J.
 PA (SHOR/) SHORT J.
 XX
 PI Callen W, Mathur EJ, Short J;
 XX
 XX WPI; 2003-182285/18.
 DR N-PSDB; ABX14886.
 XX

PT Novel DNA polymerases having increased activity and stability at
 PT increased pH and temperature, useful for DNA sequencing, amplification
 XX and incorporating non-natural nucleotides or nucleotide analogs -
 ES Claim 42; Fig 1; 81pp; English.
 XX
 CC This invention relates to a novel purified polypeptide derived from
 CC Pyrolobus fumari which has thermostable DNA polymerase activity. The
 CC protein of the invention is useful for catalyzing the formation or
 CC repair of a nucleic acid sequence, for comparing a sequence to a
 CC reference sequence and for identifying polymorphisms. The nucleic
 CC acid sequence of the invention is useful for preparing cDNA from mRNA,
 CC and in an amplification process of a double-stranded DNA molecule.
 CC The nucleotide sequence is also useful for incorporating non-natural
 CC nucleotides or its analogues into a DNA molecule, by contacting a
 CC polypeptide encoded by the nucleic acid with a DNA template in a PCR
 CC amplification reaction. The nucleotides which may be used for this are
 CC inosine, 2-aminopurine or 5-methylcytosine. A nucleic acid probe
 CC derived from the nucleic acid sequence of the invention is useful in
 CC chromosome walking procedures to identify clones containing genomic
 CC sequences located adjacent to a sequence of the gene encoding the DNA
 CC polymerases of the invention. Such methods allow the isolation of
 CC genes which encode additional proteins from the host organisms. The
 CC polymerases of the invention have increased activity and stability at
 CC increased pH and temperature, and high processivity. The present
 CC sequence represents a Pyrolobus fumarius thermostable DNA polymerase
 CC protein of the invention.

XX Sequence 803 AA;

Query Match 100.0%; Score 4188; DB 24; Length 803;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 803; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEVFTVLDSSYEVGKEPOVITWGAENGERVVLIDRSRPFYFALLAGADPKQVAQ 60
 Db 1 MTEVFTVLDSSYEVGKEPOVITWGAENGERVVLIDRSRPFYFALLAGADPKQVAQ 60
 QY 61 RIRALSREKSPITIGVEDDKRYFGRPRVLRIRTVLPEAVREYRELKQVQGVDEVLAD 120
 Db 61 RIRALSREKSPITIGVEDDKRYFGRPRVLRIRTVLPEAVREYRELKQVQGVDEVLAD 120
 QY 121 IRFAMRYLIDHDLFPFTWYVEAEPLNKGFRVDKYLAKSRPEPLGEALAPKLPDL 180
 Db 121 IRFAMRYLIDHDLFPFTWYVEAEPLNKGFRVDKYLAKSRPEPLGEALAPKLPDL 180
 QY 181 RILAFDIEVYSKQSPRPERDPVIVAVKTDGDEVLFIAEGKDKRPIRFEVVKRYD 240
 Db 181 RILAFDIEVYSKQSPRPERDPVIVAVKTDGDEVLFIAEGKDKRPIRFEVVKRYD 240
 QY 241 PDIIYGNVNHDPWYLLRRARILGKLDVTRRVGAETTSVGHVSVGRLNVDLYDA 300
 Db 241 PDIIYGNVNHDPWYLLRRARILGKLDVTRRVGAETTSVGHVSVGRLNVDLYDA 300
 QY 301 EEMPEIKISLEVAEYLGVMKKSRRVIVNWEIPDYWDPPKRPPLLIQYARDVRYTG 360
 Db 301 EEMPEIKISLEVAEYLGVMKKSRRVIVNWEIPDYWDPPKRPPLLIQYARDVRYTG 360
 QY 361 LAEKILPRAIQLSYVTGLPDQVGAMSGFRLEWYLIRAAFKKELVNRVERPEETVRG 420
 Db 361 LAEKILPRAIQLSYVTGLPDQVGAMSGFRLEWYLIRAAFKKELVNRVERPEETVRG 420
 QY 421 AIVLEPLRGVHENIAVLFSSMYPNIMIKYVNPDTLVRPGEKCGECGCEWAPVKHRR 480
 Db 421 AIVLEPLRGVHENIAVLFSSMYPNIMIKYVNPDTLVRPGEKCGECGCEWAPVKHRR 480
 QY 481 RCPDGFKTVLRLLELRKRVRAEMKKYPPDPSPEYLLDEROKALKVLANASYGMWSG 540
 Db 481 RCPDGFKTVLRLLELRKRVRAEMKKYPPDPSPEYLLDEROKALKVLANASYGMWSG 540
 QY 541 ARWYCECAKAVTAMGRHLIRTAINTARKLGKLVYGDTSLVFYDPEKVENFIKIKE 600
 Db 541 ARWYCECAKAVTAMGRHLIRTAINTARKLGKLVYGDTSLVFYDPEKVENFIKIKE 600

Qy 601 ELGFEIKLEKVKYKRLPFTPEAKRYAGLLEDRIDIVGFVAVRGDWCELAKEVQTKVEIV 660
 Db 601 ELGFEIKLEKVKYKRLPFTPEAKRYAGLLEDRIDIVGFVAVRGDWCELAKEVQTKVEIV 660
 Qy 661 LKTSVNVKAVYVYRKIVKELEBEGKVPTEKLVWKTLSKRLLEETTEAPHVAAKRLMSAG 720
 Dh 661 LKTSVNVKAVYVYRKIVKELEBEGKVPTEKLVWKTLSKRLLEETTEAPHVAAKRLMSAG 720
 Qy 721 YRVSPGDKIGVIVKGGGRISQRAWPFYFMVKDPSQIDVTYVVDHQIIPAALRILGFGIT 780
 Db 721 YRVSPGDKIGVIVKGGGRISQRAWPFYFMVKDPSQIDVTYVVDHQIIPAALRILGFGIT 780
 Qy 781 EKKLKASATGOKTLDFDLAKSK 803
 Db 781 EKKLKASATGOKTLDFDLAKSK 803

RESULT 4

ID ABG73160 standard; Protein; 803 AA.

XX ABG73160;

DT 08-APR-2003 (first entry)

XX Desulphurococcus thermostable DNA polymerase protein.

XX Enzyme; thermostable; DNA polymerase; DNA repair;

XX polymorphism identification.

XX Desulphurococcus sp.

XX Key Location/Qualifiers

FT Misc-difference 601

FT /label= xaa

FT /note= "Encoded by SCG, xaa is unknown"

XX US2002132243-A1.

XX 19-SEP-2002.

XX 06-SEP-2001; 2001US-0948369.

XX 06-AUG-1997; 97US-0907166.

XX 07-SEP-1999; 99US-0391340.

XX 06-SEP-2000; 2000US-0656309.

XX (CALL/) CALLEN W.

XX (MATH/) MATHUR E J.

XX (SHORT/) SHORT J.

XX Callen W, Mathur EJ, Short J;

XX WPI; 2003-182285/18.

XX N-PSDB; ABX14989.

XX Novel DNA polymerases having increased activity and stability at
 PT increased pH and temperature, useful for DNA sequencing, amplification
 PT and incorporating non-natural nucleotides or nucleotide analogs -

XX Claim 42; Page 46-48; 81pp; English.

XX This invention relates to a novel purified polypeptide derived from
 CC Pyrobolus fumari which has thermostable DNA polymerase activity. The
 CC protein of the invention is useful for catalyzing the formation or
 CC repair of a nucleic acid sequence, for comparing a sequence to a
 CC reference sequence and for identifying polymorphisms. The nucleic
 CC acid sequence of the invention is useful for preparing cDNA from mRNA,
 CC and in an amplification process of a double-stranded DNA molecule.
 CC The nucleotide sequence is also useful for incorporating non-natural
 CC nucleotides or its analogues into a DNA molecule, by contacting a
 CC polypeptide encoded by the nucleic acid with a DNA template in a PCR

CC amplification reaction. The nucleotides which may be used for this are
 CC inosine, 2-aminopurine or 5-methylcytosine. A nucleic acid probe
 CC derived from the nucleic acid sequence of the invention is useful in
 CC chromosome walking procedures to identify clones containing genomic
 CC sequences located adjacent to a sequence of the gene encoding the DNA
 CC polymerases of the invention. Such methods allow the isolation of
 CC genes which encode additional proteins from the host organisms. The
 CC polymerases of the invention have increased activity and stability at
 CC increased pH and temperature, and high processivity. The present
 CC sequence represents a Desulphurococcus thermostable DNA polymerase
 CC protein of the invention.

XX Sequence 803 AA;

Query Match 100.0%; Score 4188; DB 24; Length 803;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 803; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTEVFTVLDSSYEVVKGEPQVVIWGIAENGSRVVLIDRSFPPYFALLAGADPKQVAQ 60
 Db 1 MTEVFTVLDSSYEVVKGEPQVVIWGIAENGSRVVLIDRSFPPYFALLAGADPKQVAQ 60
 Qy 61 RIRALSRRPKSPIIGVEDDKRYFGPRPVLRITVLPVAVRYRRLVKNVDGVEDVLEAD 120
 Db 61 RIRALSRRPKSPIIGVEDDKRYFGPRPVLRITVLPVAVRYRRLVKNVDGVEDVLEAD 120
 Qy 121 IRFAMSYLDHDLFPFTWYRVAEPLNKGMPFRVVKVILVKSRRPEPLVGEALAPKLPDL 180
 Db 121 IRFAMSYLDHDLFPFTWYRVAEPLNKGMPFRVVKVILVKSRRPEPLVGEALAPKLPDL 180
 Qy 181 RILAPDIEVYSKQSPRPDPVIVAVTDDGDEVLFAEGKDKRKPRFVEFVVKRYD 240
 Db 181 RILAPDIEVYSKQSPRPDPVIVAVTDDGDEVLFAEGKDKRKPRFVEFVVKRYD 240
 Qy 241 PDIIVGNVNNHFDWYLLRRARILGLKDVTRVCGAEPSTTVHGVSVSGRLNVDLYDA 300
 Db 241 PDIIVGNVNNHFDWYLLRRARILGLKDVTRVCGAEPSTTVHGVSVSGRLNVDLYDA 300
 Qy 301 EEMPEIKIKSLEVAEYLVGMKKSERVIINWWEIPDYWDPPKRPPLLOQYRDDVRAATYG 360
 Db 301 EEMPEIKIKSLEVAEYLVGMKKSERVIINWWEIPDYWDPPKRPPLLOQYRDDVRAATYG 360
 Qy 361 LAEKILPFAIQLSYVTGLPDLQVGAMSVGFRLEWYLIRAFKMKELVPNRVERPESTYRG 420
 Db 361 LAEKILPFAIQLSYVTGLPDLQVGAMSVGFRLEWYLIRAFKMKELVPNRVERPESTYRG 420
 Qy 421 AIVLEPLRGVHENIAVLDESSMYPNTMIKYNVGPDTLVRPGEKCGCGWEAPEVKHRFR 480
 Db 421 AIVLEPLRGVHENIAVLDESSMYPNTMIKYNVGPDTLVRPGEKCGCGWEAPEVKHRFR 480
 Qy 481 RCPGPFKTVLERLLELRKRVRAEMKYPDPSPPEYRLDDEQKALVKLANASYGYWMSG 540
 Db 481 RCPGPFKTVLERLLELRKRVRAEMKYPDPSPPEYRLDDEQKALVKLANASYGYWMSG 540
 Qy 541 ARWYCRECAKAVTANGHRLIRTAINTARKLGLKVIYGDTSLSFVYDPEKVENFIKIKE 600
 Db 541 ARWYCRECAKAVTANGHRLIRTAINTARKLGLKVIYGDTSLSFVYDPEKVENFIKIKE 600
 Qy 601 ELGFEIKLEKVKYKRLPFTPEAKRYAGLLEDRIDIVGFVAVRGDWCELAKEVQTKVEIV 660
 Db 601 ELGFEIKLEKVKYKRLPFTPEAKRYAGLLEDRIDIVGFVAVRGDWCELAKEVQTKVEIV 660
 Qy 661 LKTSVNVKAVYVYRKIVKELEBEGKVPTEKLVWKTLSKRLLEETTEAPHVAAKRLMSAG 720
 Db 661 LKTSVNVKAVYVYRKIVKELEBEGKVPTEKLVWKTLSKRLLEETTEAPHVAAKRLMSAG 720
 Qy 721 YRVSPGDKIGVIVKGGGRISQRAWPFYFMVKDPSQIDVTYVVDHQIIPAALRILGFGIT 780
 Db 721 YRVSPGDKIGVIVKGGGRISQRAWPFYFMVKDPSQIDVTYVVDHQIIPAALRILGFGIT 780
 Qy 781 EKKLKASATGOKTLDFDLAKSK 803
 Db 781 EKKLKASATGOKTLDFDLAKSK 803

RESULT 5	
ABG73162	
ID	ABG73162 standard; Protein; 803 AA.
XX	
AC	ABG73162;
XX	
DT	08-APR-2003 (first entry)
XX	
DE	Pyrobolus thermostable DNA polymerase protein #2.
XX	
KW	Enzyme; thermostable; DNA polymerase; DNA repair;
KW	polymorphism identification.
XX	
OS	Pyrobolus fumarius.
XX	
PN	US2002132243-A1.
XX	
FD	19-SEP-2002.
XX	
FF	06-SEP-2001; 2001US-0948369.
XX	
PR	06-AUG-1997; 97US-0907166.
PR	07-SEP-1999; 99US-0391340.
PR	06-SEP-2000; 2000US-0656309.
XX	
PA	(CALL/) CALLEN W.
PA	(MATH/) MATHUR E J.
PA	(SHOR/) SHORT J.
XX	
PI	Callen W, Mathur EJ, Short J;
DR	WP1; 2003-182285/18.
DR	N-PSDB; ABX14891.
XX	
PT	Novel DNA polymerases having increased activity and stability at
PT	increased pH and temperature, useful for DNA sequencing, amplification
PT	and incorporating non-natural nucleotides or nucleotide analogs -
XX	
PS	Claim 42; Fig 1A-1E; 81pp; English.
XX	
CC	This invention relates to a novel purified polypeptide derived from
CC	Pyrobolus fumari which has thermostable DNA polymerase activity. The
CC	protein of the invention is useful for catalysing the formation or
CC	repair of a nucleic acid sequence, for comparing a sequence to a
CC	reference sequence and for identifying polymorphisms. The nucleic
CC	acid sequence of the invention is useful for preparing cDNA from mRNA,
CC	and in an amplification process of a double-stranded DNA molecule.
CC	The nucleotide sequence is also useful for incorporating non-natural
CC	nucleotides or its analogues into a DNA molecule, by contacting a
CC	polypeptide encoded by the nucleic acid with a DNA template in a PCR
CC	amplification reaction. The nucleotides which may be used for this are
CC	inosine, 2-aminopurine or 5-methylcytosine. A nucleic acid probe
CC	derived from the nucleic acid sequence of the invention is useful in
CC	chromosome walking procedures to identify clones containing genomic
CC	sequences located adjacent to a sequence of the gene encoding the DNA
CC	polymerases of the invention. Such methods allow the isolation of
CC	genes which encode additional proteins from the host organisms. The
CC	polymerases of the invention have increased activity and stability at
CC	increased pH and temperature, and high processivity. The present
CC	sequence represents a Pyrobolus fumarius thermostable DNA polymerase
CC	protein of the invention.
XX	
SQ	Sequence 803 AA;
XX	
Query Match	100.0%; Score 4188; DB 24; Length 803;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 803; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
XX	
QY	1 MTEVFTVLDSSEYVVGKEPQVIWGIAENGERVULIDRSFRPYFALLAPGADPKQVAQ 60
DB	1 MTEVFTVLDSSEYVVGKEPQVIWGIAENGERVULIDRSFRPYFALLAPGADPKQVAQ 60

```

XX PR 06-SEP-2000; 2000US-0656309.
XX PA (DIVE-) DIVERSA CORP.
XX PI Callen W, Mathur EJ, Short JM;
XX DR WPI; 2002-362247/39.
XX DR N-PSDB; AAD35188.
XX PT New thermostable polymerase useful for sequencing DNA, amplifying
XX PT double stranded DNA, or incorporating a non-natural nucleotide or a
XX PT nucleotide analog into a DNA molecule
XX PS Claim 42; Fig 11; 161pp; English.
XX CC The invention relates to thermostable DNA polymerases having high
XX CC temperature polymerase activity, such as those derived from
XX CC Pyrobolus fumaria and nucleic acid molecules encoding such polymerases.
XX CC Polymerases are useful for catalysing the formation or repair of a
XX CC nucleic acid sequence and for modifying small molecules. They are also
XX CC useful for sequencing DNA molecules, for preparing cDNA from mRNA, for
XX CC amplifying double stranded DNA molecules and for incorporating non-
XX CC natural nucleotides or nucleotide analogues into a DNA molecule. The
XX CC present sequence is Pyrobolus fumaria DNA polymerase.
XX SQ Sequence 803 AA;
    Query Match 99.4%; Score 4164; DB 23; Length 803;
    Best Local Similarity 99.3%; Pred. No. 0;
    Matches 797; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
    1 MTEVFTVLDSSVEYVKGKQVVIWGIWGAENGVRVLDIDRSPRYFVALLAPGADPKQVAQ 60
    1 MTEVFTVLDSSVEYVKGKQVVIWGIWGAENGVRVLDIDRSPRYFVALLAPGADPKQVAQ 60
    61 RIRALSRPKPIIGVEDDKRKYFGRPRRLRIRVLPEAVREYRELKVNVDGVEDVLEAD 120
    61 RIRALSRPKPIIGVEDDKRKYFGRPRRLRIRVLPEAVREYRELKVNVDGVEDVLEAD 120
    121 IRFAMRYLIDHDLFPFTWYRVEAEPLNKGFRVVKYLVKSRPEPLYGSAALPTKLPDL 180
    121 IRFAMRYLIDHDLFPFTWYRVEAEPLNKGFRVVKYLVKSRPEPLYGSAALPTKLPDL 180
    181 RILAFDIEVYKQSPRPERDPIVIAVKTDDGDEVLFIAEGCKDORPIREFVYVKRYD 240
    181 RILAFDIEVYKQSPRPERDPIVIAVKTDDGDEVLFIAEGCKDORPIREFVYVKRYD 240
    241 PDIIIVGNNHFDWPLYLRRARILGKLDVTRRVGAETTSVHGHSVPGRNLNVDLYDA 300
    241 PDIIIVGNNHFDWPLYLRRARILGKLDVTRRVGAETTSVHGHSVPGRNLNVDLYDA 300
    301 BEMPEIKISLEVAEYLVGMKSERVINWWEIPDYWDPPKQPLLLQYARDVDRATYG 360
    301 BEMPEIKISLEVAEYLVGMKSERVINWWEIPDYWDPPKQPLLLQYARDVDRATYG 360
    361 LAEKILPRAIQLSYTGTPLOVGMASVGFLEWLIIRAAFKMKELPNRVERPEETVRG 420
    361 LAEKILPRAIQLSYTGTPLOVGMASVGFLEWLIIRAAFKMKELPNRVERPEETVRG 420
    421 AIVLEPLRGVHENIAVLDPSSMYPNIMIKYNGVPTLVRPGEKCGECGCEWAEVKKRFR 480
    421 AIVLEPLRGVHENIAVLDPSSMYPNIMIKYNGVPTLVRPGEKCGECGCEWAEVKKRFR 480
    481 RCPGPFKTVLRLLELRKRVRAEMKKYPPDPSPEYRLDRLDEROKALKVLNANASYGMWSG 540
    481 RCPGPFKTVLRLLELRKRVRAEMKKYPPDPSPEYRLDRLDEROKALKVLNANASYGMWSG 540
    541 ARWYCRECAKAVTANGSHLIRTAINTARKLGKVIYGDTSLSFTVYDPEKVENFKITKE 600
    541 ARWYCRECAKAVTANGSHLIRTAINTARKLGKVIYGDTSLSFTVYDPEKVENFKITKE 600
    601 ELGFEIKLSKVYKRVLFTEAKKRYAGLLEDGRIDIVGFEAVRGDWCELAKEVQTKVVEIV 660

```

```

Db 601 ELGFEIKLSKVYKRVLFTEAKKRYAGLLEDGRIDIVGFEAVRGDWCELAKEVQTKVVEIV 660
Qy 661 LKTSVNKAVEYVYKRVIVKELEBEGKVPTEKLVIVKNTLSKRLLEYTTTEAPHVAAKRLSAG 720
Db 661 LKTSVNKAVEYVYKRVIVKELEBEGKVPTEKLVIVKNTLSKRLLEYTTTEAPHVAAKRLSAG 720
Qy 721 YRVSPGDKIGYIVKGGGRISQRAWPYFMVKDPSQIDVTYYVDHQIIPAALRILGFGIT 780
Db 721 YRVSPGDKIGYIVKGGGRISQRAWPYFMVKDPSQIDVTYYVDHQIIPAALRILGFGIT 780
Qy 781 EKKLKASATGQKTLDFDLAKSK 803
Db 781 EKKLKASATGQKTLDFDLAKSK 803

RESULT 7
ABG73161
ID ABG73161 standard; Protein; 803 AA.
XX AC ABG73161;
XX DT 08-APR-2003 (first entry)
XX DE Aquifex pyrophilus thermostable DNA polymerase protein.
XX KW Enzyme; thermostable; DNA polymerase; DNA repair;
XX KW polymorphism identification.
XX OS Aquifex pyrophilus.
XX PN US2002132243-A1.
XX PD 19-SEP-2002.
XX PF 06-SEP-2001; 2001US-0949369.
XX PR 06-AUG-1997; 97US-0907166.
XX PR 07-SEP-1999; 99US-0391340.
XX PR 06-SEP-2000; 2000US-0656309.
XX PA (CALL/) CALLEN W.
XX PA (MATHUR/) MATHUR E J.
XX PA (SHORT/) SHORT J.
XX PI Callen W, Mathur EJ, Short J;
XX PI WPI; 2003-182285/18.
XX DR N-PSDB; ABX14890.
XX CC Novel DNA polymerases having increased activity and stability at
XX CC increased pH and temperature, useful for DNA sequencing, amplification
XX CC and incorporating non-natural nucleotides or nucleotide analogs -
XX CC Claim 42; Page 49-51; 81pp; English.
XX CC This invention relates to a novel purified polypeptide derived from
XX CC Pyrobolus fumari which has thermostable DNA polymerase activity. The
XX CC protein of the invention is useful for catalysing the formation or
XX CC repair of a nucleic acid sequence, for comparing a sequence to a
XX CC reference sequence and for identifying polymorphisms. The nucleic
XX CC acid sequence of the invention is useful for preparing cDNA from mRNA,
XX CC and in an amplification process of a double-stranded DNA molecule.
XX CC The nucleotide sequence is also useful for incorporating non-natural
XX CC nucleotides or its analogues into a DNA molecule, by contacting a
XX CC polypeptide encoded by the nucleic acid with a DNA template in a PCR
XX CC amplification reaction. The nucleic acids which may be used for this are
XX CC inosine, 2-aminopurine or 5-methylcytosine. A nucleic acid probe
XX CC derived from the nucleic acid sequence of the invention is useful in
XX CC chromosome walking procedures to identify clones containing genomic
XX CC sequences located adjacent to a sequence of the gene encoding the DNA
XX CC polymerases of the invention. Such methods allow the isolation of
XX CC genes which encode additional proteins from the host organisms. The

```

CC polymerases of the invention have increased activity and stability at
 CC increased pH and temperature, and high processivity. The present
 CC sequence represents an Aquifex pyrophilus thermostable DNA polymerase
 CC protein of the invention.

XX Sequence 803 AA;

Query Match 99.4%; Score 4164; DB 24; Length 803;
 Best Local Similarity 99.3%; Pred. No. 0;
 Matches 797; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTEVFTVLDSSYEVWGKEQVLIWGIAENGERVWLDISFRFYFVALLAPGADPKQVAQ 60
 DB 1 MTEVFTVLDSSYEVWGKEQVLIWGIAENGERVWLDISFRFYFVALLAPGADPKQVAQ 60
 QY 61 RIRALSRLSPKSPFIIGVEDDKKRYFGRPRRLVIRITVLPPEAVREYRELKVNVDGVEDVLEAD 120
 DB 61 RIRALSRLSPKSPFIIGVEDDKKRYFGRPRRLVIRITVLPPEAVREYRELKVNVDGVEDVLEAD 120
 QY 121 IRFAMRYLIDHDLFPFTTWYRVEAPLENKMGFRVVKYLVKSRPELYGEALPTKLPDL 180
 DB 121 IRFAMRYLIDHDLFPFTTWYRVEAPLENKMGFRVVKYLVKSRPELYGEALPTKLPDL 180
 QY 181 RILAFDIEVYSKQSPRPDPDVIVIAVKTDDGDEVLFIAEGKDDKRPREFVEYVKRYD 240
 DB 181 RILAFDIEVYSKQSPRPDPDVIVIAVKTDDGDEVLFIAEGKDDKRPREFVEYVKRYD 240
 QY 241 PDIIYGVNNHFDWPLYLRARILGKIDVTRRVGAEPFTSVHGVSVPGRLNVDLYDYA 300
 DB 241 PDIIYGVNNHFDWPLYLRARILGKIDVTRRVGAEPFTSVHGVSVPGRLNVDLYDYA 300
 QY 301 BEMPEIKIKLEAEVAVLGVKKSERVIINWEIPDYWDPKKPELILLOYARDVRYATYG 360
 DB 301 BEMPEIKIKLEAEVAVLGVKKSERVIINWEIPDYWDPKKPELILLOYARDVRYATYG 360
 QY 361 LAEKILPAIQLSYVTGLPDQVGMSVGRLEWLIIRAAFKMKELVFNVERPEETYRG 420
 DB 361 LAEKILPAIQLSYVTGLPDQVGMSVGRLEWLIIRAAFKMKELVFNVERPEETYRG 420
 QY 421 AIVLEPLAGVHENIAVDFSSYFNIMIKYNGVPTLVRPGEKCGCGCWEAPEVGHFR 480
 DB 421 AIVLEPLAGVHENIAVDFSSYFNIMIKYNGVPTLVRPGEKCGCGCWEAPEVGHFR 480
 QY 481 RCPGFFYTVLERLLELRKRVRAEMKKYPPDPSPEYRLDEROKALKVLNAYSVMGWSG 540
 DB 481 RCPGFFYTVLERLLELRKRVRAEMKKYPPDPSPEYRLDEROKALKVLNAYSVMGWSG 540
 QY 541 ARWYCRECAKAVTAMGRHLIRTAIARGLGKVITYGTDLSLFTYDPEKVENFIKIIE 600
 DB 541 ARWYCRECAKAVTAMGRHLIRTAIARGLGKVITYGTDLSLFTYDPEKVENFIKIIE 600
 QY 601 ELGPEIKLEKVKRILFFTEAKKRYAGLLEDGRIDIVGPEAVRGWCWELAKVQKWEIV 660
 DB 601 ELGPEIKLEKVKRILFFTEAKKRYAGLLEDGRIDIVGPEAVRGWCWELAKVQKWEIV 660
 QY 661 LKTSVKNKAVEYVKIIVLEBCKVPTEKLIWIKTLSKRLSEYTTAEASHVAAKRLSAG 720
 DB 661 LKTSVKNKAVEYVKIIVLEBCKVPTEKLIWIKTLSKRLSEYTTAEASHVAAKRLSAG 720
 QY 721 YRVSPGDKIGYIVVKGGRISORAMPYFMWKDPSQIDVTYVVDHQIIPAAKLRIILGYGIT 780
 DB 721 YRVSPGDKIGYIVVKGGRISORAMPYFMWKDPSQIDVTYVVDHQIIPAAKLRIILGYGIT 780
 QY 781 EKKLKASATGQKTLDFDLAKSK 803
 DB 781 EKKLKASATGQKTLDFDLAKSK 803

RESULT 8

ABG73163

ID ABG73163 standard; Protein; 803 AA.

XX

AC ABG73163;

XX

DT

XX

DE

XX

KW

KW

XX

OS

XX

PN

XX

PD

XX

PF

XX

XX

PR

PR

PR

XX

PA

PA

XX

PI

XX

DR

XX

XX

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Db 121 IRFAMKYLIDHDLFPFTWYRVEAEPLNKGFRVDKVLVKSRPEPLYGEALPTKLPDL 180
 QY 181 RILAFDIEVYKQSGSRPRDPVIVIAVKTDDGDEVLFIAEGKDDKRPPIREFVYVKRYD 240
 Db 181 RILAFDIEVYKQSGSRPRDPVIVIAVKTDDGDEVLFIAEGKDDKRPPIREFVYVKRYD 240
 QY 241 PDIIVGNVNNHFDWPLYLRRARILGIKLDVTRRVGABPTTSVHGHVSVPGLNVDLYDA 300
 Db 241 PDIIVGNVNNHFDWPLYLRRARILGIKLDVTRRVGABPTTSVHGHVSVPGLNVDLYDA 300
 QY 301 EEMPEIKISLEEVAEYLVGKMSERVIIINWBEIPDYWDDPKKPLLIQVARDVRYATG 360
 Db 301 EEMPEIKISLEEVAEYLVGKMSERVIIINWBEIPDYWDDPKKPLLIQVARDVRYATG 360
 QY 361 LAEKILPFAIQLSYVTGLPQVGMSVGFLEWYLIRAAFAFMKELVFNVRVERPEEYVRG 420
 Db 361 LAEKILPFAIQLSYVTGLPQVGMSVGFLEWYLIRAAFAFMKELVFNVRVERPEEYVRG 420
 QY 421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYNNVGPDTLVRPGECGCGCEWAEPEVKHRFR 480
 Db 421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYNNVGPDTLVRPGECGCGCEWAEPEVKHRFR 480
 QY 481 RCPDPPFTVLERLLELRKRVRAEMKKYPPDPSPEYRLDERQKALKVLANASYGMWSG 540
 Db 481 RCPDPPFTVLERLLELRKRVRAEMKKYPPDPSPEYRLDERQKALKVLANASYGMWSG 540
 QY 541 ARWYCRECAKAVTAWGRHLIRTAIRAKLGLKVYIGDTSIFVTYDPEKVENFIKIE 600
 Db 541 ARWYCRECAKAVTAWGRHLIRTAIRAKLGLKVYIGDTSIFVTYDPEKVENFIKIE 600
 QY 601 ELGFEIKLEKVKYKRLFFTEAKRYAGLLEDGRIDIVGFEAVRGDWCELAKEVQTKWEIV 660
 Db 601 ELGFEIKLEKVKYKRLFFTEAKRYAGLLEDGRIDIVGFEAVRGDWCELAKEVQTKWEIV 660
 QY 661 LKTSVNVKAVEYVRKIVKELEBKVPIEKLVIWKTLSKRLBEYTTTEAPHVVAAKRMSAG 720
 Db 661 LKTSVNVKAVEYVRKIVKELEBKVPIEKLVIWKTLSKRLBEYTTTEAPHVVAAKRMSAG 720
 QY 721 YRVSPGDKIGYIVVKGGRISORAWPYFMWKPDSQIDVTYVVDHQIIPAAALRILGYFGIT 780
 Db 721 YRVSPGDKIGYIVVKGGRISORAWPYFMWKPDSQIDVTYVVDHQIIPAAALRILGYFGIT 780
 QY 781 EKKLKASATGQKTLDFDLAKSK 803
 Db 781 EKKLKASATGQKTLDFDLAKSK 803

RESULT 9

AAE22120
 ID AAE22120 standard; Protein; 799 AA.

XX AC AAE22120;

XX DT 25-JUL-2002 (first entry)

XX DE Pyrolobus fumaria DNA polymerase consensus sequence.

XX KW DNA polymerase; thermostable; enzyme.

XX OS Pyrolobus fumaria.

XX PN WO200220735-A2.

XX PD 14-MAR-2002.

XX PF 06-SEP-2001; 2001WO-US28007.

XX PR 06-SEP-2000; 2000US-0656309.

XX PA (DIVE-) DIVERSA CORP.

XX PI Callen W, Mathur EJ, Short JM;

DR MF1; 2002-362247/39.

XX New thermostable polymerase useful for sequencing DNA, amplifying
 PT double stranded DNA, or incorporating a non-natural nucleotide or a
 PT nucleotide analog into a DNA molecule -

XX Disclosure; Fig 11; 161pp; English.

PS The invention relates to thermostable DNA polymerases having high
 XX temperature polymerase activity, such as those derived from
 CC Pyrolobus fumaria and nucleic acid molecules encoding such polymerases.
 CC Polymerases are useful for catalyzing the formation or repair of a
 CC nucleic acid sequence and for modifying small molecules. They are also
 CC useful for sequencing DNA molecules, for preparing cDNA from mRNA, for
 CC amplifying double stranded DNA molecules and for incorporating non-
 CC natural nucleotides or nucleotide analogues into a DNA molecule. The
 CC present sequence is Pyrolobus fumaria DNA polymerase consensus sequence.

XX SQ Sequence 799 AA;

Query Match 98.4%; Score 4122; DB 23; Length 799;
 Best Local Similarity 99.4%; Pred. No. 0;
 Matches 798; Conservative 1; Mismatches 0; Indels 4; Gaps 4;

QY 1 MTEVVTVDSSYEVYVKGSPQVLIINGIAENGSRVLIIDRSFRPYFVALLAPADPKQVAQ 60
 Db 1 MTEVVTVDSSYEVYVKGSPQVLIINGIAENGSRVLIIDRSFRPYFVALLAPADPKQVAQ 60
 QY 61 RIBALSRKSPITIGVEDDKKYGRRVRVLRITVLPVAVREYRELKVVGVGVLEAD 120
 Db 61 RIBALSRKSPITIGVEDDKKYGRRVRVLRITVLPVAVREYRELKVVGVGVLEAD 120
 QY 121 IRFAMKYLIDHDLFPFTWYRVEAEPLNKGFRVDKVLVKSRPEPLYGEALPTKLPDL 180
 Db 121 IRFAMKYLIDHDLFPFTWYRVEAEPLNKGFRVDKVLVKSRPEPLYGEALPTKLPDL 180
 QY 181 RILAFDIEVYKQSGSRPRDPVIVIAVKTDDGDEVLFIAEGKDDKRPPIREFVYVKRYD 240
 Db 181 RILAFDIEVYKQSGSRPRDPVIVIAVKTDDGDEVLFIAEGKDDKRPPIREFVYVKRYD 240
 QY 241 PDIIVGNVNNHFDWPLYLRRARILGIKLDVTRRVGABPTTSVHGHVSVPGLNVDLYDA 300
 Db 241 PDIIVGNVNNHFDWPLYLRRARILGIKLDVTRRVGABPTTSVHGHVSVPGLNVDLYDA 300
 QY 301 EEMPEIKISLEEVAEYLVGKMSERVIIINWBEIPDYWDDPKKPLLIQVARDVRYATG 360
 Db 301 EEMPEIKISLEEVAEYLVGKMSERVIIINWBEIPDYWDDPKKPLLIQVARDVRYATG 360
 QY 361 LAEKILPFAIQLSYVTGLPQVGMSVGFLEWYLIRAAFAFMKELVFNVRVERPEEYVRG 420
 Db 361 LAEKILPFAIQLSYVTGLPQVGMSVGFLEWYLIRAAFAFMKELVFNVRVERPEEYVRG 420
 QY 421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYNNVGPDTLVRPGECGCGCEWAEPEVKHRFR 480
 Db 421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYNNVGPDTLVRPGECGCGCEWAEPEVKHRFR 480
 QY 481 RCPDPPFTVLERLLELRKRVRAEMKKYPPDPSPEYRLDERQKALKVLANASYGMWSG 540
 Db 481 RCPDPPFTVLERLLELRKRVRAEMKKYPPDPSPEYRLDERQKALKVLANASYGMWSG 540
 QY 541 ARWYCRECAKAVTAWGRHLIRTAIRAKLGLKVYIGDTSIFVTYDPEKVENFIKIE 600
 Db 541 ARWYCRECAKAVTAWGRHLIRTAIRAKLGLKVYIGDTSIFVTYDPEKVENFIKIE 600
 QY 601 ELGFEIKLEKVKYKRLFFTEAKRYAGLLEDGRIDIVGFEAVRGDWCELAKEVQTKWEIV 660
 Db 601 ELGFEIKLEKVKYKRLFFTEAKRYAGLLEDGRIDIVGFEAVRGDWCELAKEVQTKWEIV 660
 QY 661 LKTSVNVKAVEYVRKIVKELEBKVPIEKLVIWKTLSKRLBEYTTTEAPHVVAAKRMSAG 720
 Db 661 LKTSVNVKAVEYVRKIVKELEBKVPIEKLVIWKTLSKRLBEYTTTEAPHVVAAKRMSAG 720
 QY 721 YRVSPGDKIGYIVVKGGRISORAWPYFMWKPDSQIDVTYVVDHQIIPAAALRILGYFGIT 780

```

Db      717 YRVSPGDIGVIVKGGRIISQRAWPFVWVKDPSQIDVTYVDHQIIPAAIRILGYFGIT 776
QY      781 EKLLKASATGQKTLFDLFLAKSK 803
Db      777 EKLLKASATGQKTLFDLFLAKSK 799

RESULT 10
AAR92523
ID      AAR92523 standard; Protein; 803 AA.
AC      AAR92523;
XX      12-JUL-1996 (first entry)
DT      Pyrodictum occultum 803 amino acid DNA polymerase gene.
DE      DNA pol; base pair; thermally stable; exonuclease activity.
KW      Pyrodictum occultum DSM2709.
OS      Pyrodictum occultum DSM2709.
XX      JP07327684-A.
FN      19-DEC-1995.
PD      09-JUN-1994; 94JP-0150591.
PF      09-JUN-1994; 94JP-0150591.
PR      (TAKI) TAKARA SHUZO CO LTD.
PA      WPI; 1996-072342/08.
DR      DNA encoding DNA polymerase - useful for prodn. of thermally stable
PT      enzyme
PT      Claim 1; Page 11-13; 23pp; Japanese.
PS      AAR92522 and AAR92523 are 914 amino acid and 803 amino acid DNA
CC      polymerases, respectively. They are derived from Pyrodictum
CC      occultum and are thermostable DNA polymerases. Also disclosed
CC      are DNA polymerase genes which hybridize to the above genes.
XX      Sequence 803 AA;
SQ      Query Match 72.4%; Score 3033.5; DB 17; Length 803;
        Best Local Similarity 70.9%; Pred. No. 1e-276;
        Matches 571; Conservative 106; Mismatches 121; Indels 7; Gaps 5;

QY      1 MTEVV-FTVLDSYVVGKRFQVITWGAENGRIIVLIDRSFRPFYVALLAPGADP--KQ 57
Db      1 MTETIEFVLDSYVEILKCEPVVILWGLTLDGKRVLLDHRFPFYVALLARGVEDMVEE 60
QY      58 VAQRIRALSRPKSPIIGVEDDKRYFGPRPRVLRITVLPVAVREYELVKNVGDVVL 117
Db      61 TAASIRRLSVKSPSIIIDAKPLDKRYFGPRKAVKITTMPESVSHYEAVKKIEGVDSL 120
QY      118 EADIRFAMKYLIDHLPFFTVRYEAEPLEKNMGFRVVKVLYKSRPEPLYGEALAPTKL 177
Db      121 EADIRFAMKYLIDHLPFFTVRYEAPLEKNMGFRVVKVLYKSRPEPLYGEALAPTKL 178
QY      178 PLRLILAFDIEVYKQSGSPRPERDPVIVIAVKTDDGDEVLFIAEGKDDRKPIREFVEYVK 237
Db      179 PPMRLVAFDIEVYKQSGSPRPERDPVIVIAVKTDDGDEVLFIAEGKDDRKPIREFVEYVK 238
QY      238 RYDPLIIVGNNHFDWYLARRALIGIKLDTVRVGAETTSVHGHSVPGELNVLDY 297
Db      239 AFDPLIIVGNNHFDWYLARRALIGIKLDTVRVGAETTSVHGHSVPGELNVLDY 298
QY      298 DYAEEMPIKIKSLAEVAEYLGWVKSRVIVNWEIPDYWDPPKPKPLQLQVARDVRA 357
Db      299 DYAEEMPIKIKSLAEVAEYLGWVKSRVIVNWEIPDYWDPPKPKPLQLQVARDVRA 358

```

```

QY      358 TYGLAEKILPPAIQLSVTGLDQVAMSVGRLEWYLIRAAFKMGKGLVNNRVERPSET 417
Db      359 TYGLAEKMLPPAIQLSVTGLDQVAMSVGRLEWYLIRAAFKMGKGLVNNRVERPSET 418
QY      418 YRGATVLEPLRGVHENIAVLDFSSMYPNIMIKYNGDPTLVRFGEKGC-CCWEAPEYK 476
Db      419 YKGAVVLKPLAGVHENVVLDSSMYPNIMIKYNGDPTLVRFGEKGC-CCWEAPEYK 478
QY      477 HRFRCPPGPFKTVLERLELRKRVRAEMKKYPPDSPYEYLLDEROKALKVLANASYGM 536
Db      479 HRFRCPPGPFKTVLERLELRKRVRAEMKKYPPDSPYEYLLDEROKALKVLANASYGM 538
QY      537 GWSGARWYCREKAVTANGSHLIRTAINTARKLGLKVIYDGLTSLVPTDPRKVENFIK 596
Db      539 GWSGARWYCREKAVTANGSHLIRTAINTARKLGLKVIYDGLTSLVPTDPRKVENFIK 598
QY      597 IIKBELGFEIKLBKVIYKRLFFTEAKRYAGLLEGRIDIVGFEAVRGDMCELAKYQTKV 656
Db      599 FVEKELGFEIKLBKVIYKRLFFTEAKRYAGLLEGRIDIVGFEAVRGDMCELAKYQTKV 658
QY      657 VEIVLKTSEVNAVEYKRVKIVKLEBKVPKIEKLVWKTLSKLESEYTTTAPHVAAKRM 716
Db      659 AEIVLNTGNDVKAIYSIREVVKQLEKGVPTIKLIIWKTLSKLESEYTTTAPHVAAKRM 718
QY      717 LSAGYRVSPGDKIGVIVKGGRIISQRAWPFVWVKDPSQIDVTYVDHQIIPAAIRILGY 776
Db      719 KEAGYEVSPGDKIGVIVKGGRIISQRAWPFVWVKDPSQIDVTYVDHQIIPAAIRILGY 777
QY      777 FGVTEKQLKAAATVQRSLEDFEFASK 802
Db      778 FGVTEKQLKAAATVQRSLEDFEFASK 802

```

RESULT 11

AAR60883
ID AAR60883 standard; Protein; 803 AA.

XX AAR60883;
AC AAR60883;
XX 25-MAR-2003 (updated)
DT 19-JUL-1995 (first entry)
DE P. occultum DNA polymerase.

XX Polymerase chain reaction; PCR; amplify; primer; probe;
KW Pyrodictum occultum; DNA polymerase; P. abyssal; condensation;
KW DNA sequencing; nick translation; reverse transcription;
KW heat stability; heat resistant polymerase; DNA helix stabiliser.

XX Pyrodictum occultum.
XX BP624641-A2.

XX 17-NOV-1994.
XX 02-MAY-1994; 94BP-0106811.

XX 14-MAY-1993; 93US-0062368.
XX (HOFF) HOFFMANN LA ROCHE & CO AG F.

XX Gelfand DH, Wang AM;
XX WPI; 1994-350781/44.

XX N-PSDB; AAQ73844.
XX New heat stable nucleic acid polymerase from Pyrodictum species

PT - and relates DNA, vectors and transformed hosts, resistant to
PT denaturation at 100 deg.C, useful for nucleic acid amplification,
PT in DNA sequencing etc.
XX Claim 10; Page 33-35; 45pp; English.

XX This sequence represents the Pyrodicticum occultum DNA polymerase gene.
 CC The P. occultum DNA polymerase gene shows regions of homology to the
 CC P. abyssi DNA polymerase genes, and the primers given in AAQ73824-25
 CC are based on these regions. The amplified DNA encodes this DNA
 CC polymerase which catalyses condensation of dNTPs to form complementary
 CC strand DNA. This polymerase may be used in nucleic acid amplification,
 CC partic. PCR, but also for DNA sequencing, nick translation and reverse
 CC transcription. The polymerase has better heat stability than known
 CC heat resistant polymerases. Temperatures of 100 deg. C can be
 CC tolerated, facilitating strand separation and elimination of the
 CC need for DNA helix stabilisers.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX SQ Sequence 803 AA;
 Query Match 72.3%; Score 3029.5; DB 15; Length 803;
 Best Local Similarity 70.8%; Pred. No. 3e-276;
 Matches 570; Conservative 106; Mismatches 122; Indels 7; Gaps 5;
 QY 1 MTEVV-FTVLDSSYEVGKPEQVIINGIAENGERVVLIDRSFRPFYFALLAGADP--KQ 57
 DB 1 MTEIEFVLDDSSYEILGKPEVILMGITLDGKRVLLDHRPFYFALLARGYEDMVEE 60
 QY 58 VAQIRALSRPKSPIIGVEDDKRYFGRPRRLVIRTVLPEAVREYRELKVNQVGVEDVL 117
 DB 61 IAASTRRLSVVKSPIDAKPLDKRYFGRPRKAVKITTIPESVHRVREAVKKIEGVEDSL 120
 QY 118 EADIRFARMYLIHDHLPFTWYVREAEPLNKMGRVVDKVLVKSRLPELYGEALAPTKL 177
 DB 121 EADIRFARMYLIHDHLPFTWYVREAEPLNKMGRVVDKVLVKSRLPELYGEALAPTKL 178
 QY 178 PDLARILADIEVSKQSPRPDPVIVIAVKTGDGDEVLFIAEGKDDRKPIREFEYVK 237
 DB 179 PPMRLVAFDIEVSRGSPNAPDPVILVSLRDSGKERLIEAGHDDRRVREFVEYVR 238
 QY 238 RYDDIIVGNNHFDPIYLRRLRILGILKLDVTRVGAEPDTSVHGHSVPGRLNVDLY 297
 DB 239 AFDEPIIVGNSNHFDPYLMERARRLGINDVTRVGAEPDTSVHGHSVQGSRLNVDLY 298
 QY 298 DYAEEMPEIKTSLSEVAEYLGVMKSRVILNWEIPDYWDKPKRPLLLQVARDVRA 357
 DB 299 DYAEEMPEIKMTLEVAEYLGVMKSRVILNWEIPDYWDKPKRPLLLQVARDVRA 358
 QY 358 TYGLAEKILPFAIQLSYVTGLPDLQVGMVSGFRLWYLIRAPFMKELVNNRVERPEET 417
 DB 359 TYGLAEKMLPFAIQLSYVTGLPDLQVGMVSGFRLWYLIRAPFMKELVNNRVERPEET 418
 QY 418 YRGALVLEPLRGVHNIADLVDFSSMYPMIMIKYNGVPTLVPGKCGEC-GCWEAPEVK 476
 DB 419 YKGAVLKLKGVHNVVLDVDFSSMYPMIMIKYNGVPTLVPGKCGEC-GCWEAPEVK 478
 QY 477 HRFRCPPGFFVTLELLELKRVRASMKKYPDPSPRYLLDERQALKVLNANASYGM 536
 DB 479 HRFRCPPGFFVTLELLELKRVRASMKKYPDPSPRYLLDERQALKVLNANASYGM 538
 QY 537 GWSGARWYCRCAKATWAGRHILINTAINIARKLGLKVIYGDTSLSFVYDPEKVENFIK 596
 DB 539 GWSHARWYCRCAKATWAGRHILINTAINIARKLGLKVIYGDTSLSFVYDPEKVENFIK 598
 QY 597 IIKELGFEIKLVKYLFPTEAKRYAGLLEDGRIDIVGFEAVRGDMCELAKEVQTKV 656
 DB 599 FVEKSLGFEIKDKTKYKVFTEAKRYAGLLEDGRIDIVGFEAVRGDMCELAKEVQTKV 658
 QY 657 VBIVLKTEVNKAQVYRKIVKLEEGKVPTEKLVIMWTKLSKRLPEYVTEAPHVAAKRM 716
 DB 659 ABIVLNTGNVDAISYIREVINKQLREGKVPITKLIWKILSKRIEYEHDAAPHVAAKRM 718
 QY 717 LSAGYVSPGDKIGYVIVKGGGRIISQRAWPFVFWKDPSCQIDVTYVYDHIIPAALRIIGY 776
 DB 719 KEAGYVSPGDKIGYVIVKGGGRIISQRAWPFVFWKDPSCQIDVTYVYDHIIPAALRIIGY 777
 QY 777 FGITEKKUKASATGQKTLDFDLAKK 801

DB 778 FGITEKKUKASATGQKTLDFDLAKK 802
 RESULT 12
 AAR60882
 ID AAR60882 standard; Protein; 803 AA.
 XX AAR60882;
 AC AAR60882;
 DT 25-MAR-2003 (updated)
 DT 19-JUL-1995 (first entry)
 XX
 DE P. abyssi DNA polymerase.
 XX
 KW Polymerase chain reaction; PCR; amplify; primer; probe;
 KW Pyrodicticum occultum; DNA polymerase; P. abyssi; condensation;
 KW DNA sequencing; nick translation; reverse transcription;
 KW heat stability; heat resistant polymerase; DNA helix stabiliser.
 XX
 OS Pyrodicticum abyssi.
 XX
 PN EP624641-A2.
 XX
 PD 17-NOV-1994.
 XX
 PP 02-MAY-1994; 94EP-0106811.
 XX
 PR 14-MAY-1993; 93US-0062368.
 XX
 PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
 XX
 PI Gelfand DH, Wang AM;
 XX
 XX WPI; 1994-350781/44.
 DR N-PSDB; AAQ73843.
 DR
 XX New heat stable nucleic acid polymerase from Pyrodicticum species
 PT - and relates DNA, vectors and transformed hosts, resistant to
 PT denaturation at 100 deg.C, useful for nucleic acid amplification,
 PT in DNA sequencing etc.
 XX
 PS Claim 8; Page 28-31; 45pp; English.
 XX
 CC This sequence represents the Pyrodicticum abyssi DNA polymerase gene.
 CC The P. abyssi DNA polymerase gene shows regions of homology to the
 CC P. occultum DNA polymerase gene, and the primers given in AAQ73824-25
 CC are based on these regions. The amplified DNA encodes this DNA
 CC polymerase which catalyses condensation of dNTPs to form complementary
 CC strand DNA. This polymerase may be used in nucleic acid amplification,
 CC partic. PCR, but also for DNA sequencing, nick translation and reverse
 CC transcription. The polymerase has better heat stability than known
 CC heat resistant polymerases. Temperatures of 100 deg. C can be
 CC tolerated, facilitating strand separation and elimination of the
 CC need for DNA helix stabilisers.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX SQ Sequence 803 AA;
 Query Match 70.8%; Score 2963.5; DB 15; Length 803;
 Best Local Similarity 68.4%; Pred. No. 4.2e-270;
 Matches 551; Conservative 118; Mismatches 129; Indels 7; Gaps 5;
 QY 1 MTEVV-FTVLDSSYEVGKPEQVIINGIAENGERVVLIDRSFRPFYFALLAGADP--KQ 57
 DB 1 MTEIEFVLDDSSYEILGKPEVILMGITLDGKRVLLDHRPFYFALLARGYEDMVEE 60
 QY 58 VAQIRALSRPKSPIIGVEDDKRYFGRPRRLVIRTVLPEAVREYRELKVNQVGVEDVL 117
 DB 61 IAASTRRLSVVKSPIDAKPLDKRYFGRPRKAVKITTIPESVHRVREAVKKIEGVEDSL 120
 QY 118 EADIRFARMYLIHDHLPFTWYVREAEPLNKMGRVVDKVLVKSRLPELYGEALAPTKL 177

Db 121 EADIRAMRYLIDKLLPYTAYVRAENAGSPGFRVDSVYTVIVDPBPPI--ADITSIDI 178
 QY 178 PDRLILAFIEVYKSGSPRPDPVIVIAVKTDDGDEVLFIAEGKDRKPRFEPVEYVK 237
 Db 179 FENRVLAFDIEVYKSGSPRPDPVILIIKDKSGNEKLEANNYDDRNVLREFIEYIR 238
 QY 238 RYDPDIIIVGNVNNHFWPVLRLRABTLGLDVTTRVGAEPPTSVHGHVSVPGRLNVDLY 297
 Db 239 SFDPDIIIVGNVNNHFWPVLRLRABTLGLDVTTRVGAEPPTSVHGHVSVPGRLNVDLY 298
 QY 298 DYAEEMPEIKSLBEEVAEYLGVMKKSERVIINWWEIPDYWDPKRKLLOVARDVRA 357
 Db 299 NYVEEMHEIKVTLBEEVAEYLGVMKKSERVLIEWRIPDYWDDEKRLKRYALDDVRA 358
 QY 358 TYGLAEKILFPAIQLSVVTGLPDDQVGMVSGRLEWYLIRAAFKKELVPRVERPEET 417
 Db 359 TYGLAEKILFPAIQLSVVTGLPDDQVGMVSGRLEWYLIRAAFKKELVPRVERPEET 418
 QY 418 YRGAIVLEPLRGVHENIAVLDFSSMYPNIMKYNVGPDTLVRPGEKCGS--CGCWEAPEVK 476
 Db 419 YKGAIVLPLRGVHENIAVLDFSSMYPNIMKYNVGPDTLVRPGEKCGS--CGCWEAPEVK 478
 QY 477 HRFRRCPPGFFKTVLERLLELRKVRRAEMKKYPPDSPSEYKLLDERQKALKVLANASYGM 536
 Db 479 HMFRRSPGFFKTVLERLLELRKVRRAEMKKYPPDSPSEYKLLDERQKALKVLANASYGM 538
 QY 537 GWSGARWYCRCAKAVTANGRHILRTAINARKLGLKVIYVGTDSLSFTVDPKVENFIK 596
 Db 539 GWVHARWYCRCAKAVTANGRHILRTAINARKLGLKVIYVGTDSLSFTVDPKVENFIK 598
 QY 597 IIXEELGFEKLEKVKYKRLPFTFAKRYAGLLEDGRDIVGFAVRGDMCELAKVQTKV 656
 Db 599 FVEKQLGFEKIDKVIYKRVFFTEAKRYVGLLEDGRMDIVGFAVRGDMCELAKVQTKV 658
 QY 657 VEIVLKTSEVNAKVEYVYKVIKLEBEGKVPTEKLVINWTLKSLLEEYTTAPHVVAARM 716
 Db 659 AEILKTGDINRAISYIREVVVKLRREGKIPITKLVINWTLKSLLEEYTTAPHVVAARM 718
 QY 717 LSAGYRVSPEGKIGYIVVKGGRISQRAWPMYKDPQSDIVTVYVDHQLIPALRLTGY 776
 Db 719 KEAGYDVAAGDKIGYIVVKGGRISQRAWPMYKDPQSDIVTVYVDHQLIPALRLTGY 777
 QY 777 FGITEKKLSATGKTLDFDLAKK 801
 Db 778 FGVTEKQLKAASSGHRSLDFDLAKK 802

RESULT 13

AA19530
 ID AAB19530 standard; Protein; 785 AA.

XX AA19530;

XX 09-JAN-2001 (first entry)

XX Pyrobaculum islandicum DNA polymerase.

XX DNA polymerase; 3'-5' exonuclease; thermostable; DNA amplification;
 XX DNA sequencing.

XX Pyrobaculum islandicum.

XX WO200053772-A1.

XX 14-SEP-2000.

XX 04-MAR-2000; 2000WO-EP01917.

XX 06-MAR-1999; 99EP-0104539.

XX 02-SEP-1999; 99EP-0117245.

XX (HOFF) ROCHE DIAGNOSTICS GMBH.

PI Kaehler M, Frey B, Sobek H, Antranikian G;
 XX WPI; 2000-587439/55.
 DR N-PSDB; AA50678.
 XX Novel DNA polymerase of Pyrobaculum islandicum used for determining DNA
 PT sequences, comprising six conserved motifs indicative of family B DNA
 PT polymerases and possessing exonuclease activity -
 XX Disclosure; Fig 1/2; 38pp; English.
 XX The present sequence is that of the novel DNA polymerase of the
 CC anaerobic hyperthermophilic Archaeon, Pyrobaculum islandicum DSMZ4184.
 CC The sequence was deduced from a DNA polymerase gene (see AA50678)
 CC isolated from P. islandicum chromosomal DNA. The enzyme includes 6
 CC conserved motifs, indicative of a family B DNA polymerase. A method
 CC for producing recombinant DNA polymerase using Escherichia coli BL21
 CC cells transformed by a vector including the P. islandicum gene is
 CC claimed. The P. islandicum DNA polymerase is thermostable, having a
 CC half-life of 30-40 min at 100 degree C and over 10 hr at 90 degree C.
 CC It can be used for in vitro amplification of DNA fragments and for
 CC DNA sequencing (claimed). The enzyme also has 3'-5' proofreading
 CC exonuclease activity, providing a substantially lower base
 CC incorporation error rate than non-proofreading DNA polymerases.
 XX Sequence 785 AA;

Query Match 50.8%; Score 2129.5; DB 21; Length 785;
 Best Local Similarity 53.4%; Pred. No. 2.2e-191;
 Matches 429; Conservative 141; Mismatches 195; Indels 39; Gaps 13;
 QY 9 LDSSVEVVGKBPQVVIWGIANGSERVWILIDRSFRFFY---ALLAPGADPKQVAQRIRAL 65
 Db 8 LITVAVGVSVPEIRIFGLSSGERVWILIDRSFKFYVDCVCEPAA-----LKTAL 60
 QY 66 SRPKSPIIGVEDDKKIFGFRPRVLRITVLPFAVREVELVKNVGVDEVLDAIRFAM 125
 Db 61 SR-VAPIDVQIVERRFLGRSKFLKVIKIPEDYKLEAAMSIPRVSGVYEADIRFYM 119
 QY 126 RYLIDHDLPPFWYVRAEPLNKGFRVDK--YLVKSRPRLYXG--EALAPTLPDLR 181
 Db 120 RYMDMGVVPVCSNVAEEV-----GRLGGITVVSQ---WGLDEGFPFS---LK 166
 QY 182 ILAPDIEVYKSGSPRPDPVIVIAVKTDDGDEVLFIAEGKDRKPRFEPVEYKRYDP 241
 Db 167 VMAFDIEVYNERGSPDPIDRDPVVMIAIKINDGHEEVFEASGKDRGVRAFVDFIRSYDP 226
 QY 242 DIIVGNVNNHFWPVLRLRABTLGLDVTTRVGAEPPTSVHGHVSVPGRLNVDLYDAE 301
 Db 227 DVIIVGNVNGFDPVYLVERAKAVGPELVKVD--RLSNPPQOSVYGHWSIVGRANVDLYNIVE 285
 QY 302 EMPEIKISLEVAEYLGVMKKSERVIINWWEIPDYWDPKRKLLOVARDVRAVYGL 361
 Db 286 EFPEIKLTLDRVAEYFGVKNKEERVLIPCHKIYEYKDPNKRPLKRYLDDVRSFTGL 345
 QY 362 AEKILFPAIQLSVVTGLPDDQVGMVSGRLEWYLIRAAFKKELVPRVERPEETVYGA 421
 Db 346 ADKILFPLIQLSSVSGLEPLDQVAAASVGNKRVENMLRYAYRLGEVAPNREEREYEPYKA 405
 QY 422 IVLEPLRGVHENIAVLDFSSMYPNIMKYNVGPDTLVRPGEKCGCWEAPEVKHFR 481
 Db 406 IVLEPKGMVYEDVLVDFSSMYPNIMKYNLSPTLYLSEFPGEPPPPGPNVAPVSGHFRFR 465
 QY 482 CPFGFFKTVLERLLELRKVRRAEMKKYPPDSPSEYKLLDERQKALKVLANASYGMWSGA 541
 Db 466 SPFGFVQVLSLVELRKAVREAAKYPDPSPFEKILDERQKALKVMAIYGLNWSGA 525
 QY 542 RMYCREKAKAVTANGRHILRTAINARKLGLKVIYVGTDSLSFTVDPKVENFIKKEE 601
 Db 526 RMYCREVAESVTAFAAILKDVIEQARRLGIIVVYVGTDSLSFTVDPKVENFIKKEE 584
 QY 602 LGPEIKLEKVKYKRLPFTFAKRYAGLLEDGRDIVGFAVRGDMCELAKVQTKVIVL 661

XX 06-AUG-1998; 98DE-1035653.
PR 07-SEP-1998; 98DE-1040771.
PR 18-JUN-1999; 99EP-0111795.
XX (LION-) LION BIOSCIENCE AG.
XX
PI Kilger C, Kober I, Voss H, Moeckel G;
XX WPI; 2000-195576/17.
DR
XX Thermostable in vitro polymerase complex for template-dependent
PT elongation of nucleic acids
PT
XX Disclosure; Page 167-170; 233pp; German.
PS
XX This invention describes a novel thermostable in vitro complex (1) for
CC template-dependent elongation of nucleic acids (NA) comprising a
CC thermostable 'staple' protein and an elongation protein. The thermostable
CC in vitro complex is useful for template-dependent elongation of NA, e.g.
CC for amplification or reverse transcription. This is useful for sequencing
CC NA by the polymerase chain reaction or reverse transcriptase-PCR. The
CC complex can be used to mark NA. This sequence represents the Archaeoglobus
CC fulgidus AF0497 protein which is used to illustrate the method of the
CC invention.
XX
SQ Sequence 781 AA;
Query Match 37.8%; Score 1581; DB 21; Length 781;
Best Local Similarity 42.5%; Pred. No. 1.3e-139;
Matches 350; Conservative 133; Mismatches 273; Indels 68; Gaps 18;
QY 1 MTEVFTVLDSSYEVNKEPQVILNGIAENGKRVILDRSPRPYFALLAPGADPKQVAQ 60
DB 1 MERVGLIDADYETIGGKAVRLMCKDDQG-IFVAYDYNFDPYFVI---GVDEDILKN 56
QY 61 RIRALSPKSPILIGVEDDKRYKFGPRVLRIRVLPVAVREYVELKVNVDGVEDVLEAD 120
DB 57 AATSTRREVILKSKFEKQKLTGLREVEGYIVYAAHHQHVPLKADYLSQFG---DVREAD 113
QY 121 IRFAMRVLIIDHLPFFFTWYRVEAPLENKMG-----FRVDKVLVKSRRPELYGEALAPT 175
DB 114 IPFAYRYLLDKDLACMDGIAEGE---KQGGVIRSYKIEKV-----ERIPRM 157
QY 176 KLPDLRIILAFIEVYKSGSPRSDPVIIVIAVKTDDGDEVLFIAEGKDDKPKPIREVEY 235
DB 158 EFPKLMVDFCEMLSPGMPPEKDPPIIVISVKTNDDEIILTG---DERKIISDFVKL 214
QY 236 VKRYDPDIIIVGNNHFDMPYLLRRARILGKLDVTRVSGAEPSTSVH---GHVSVPGRIN 293
DB 215 IKSYPDIIIVGYNQDAFDPYLRKKAERNIPLD-----VGRDGSNVVPRGGRPKITGELN 270
QY 294 VLDYDABEMPEIKISLEAEVLYGVNKKSERVLIINWETIPDYWDDPKKPLLLQYARD 353
DB 271 VLDYDIAMRISDIKIKLENVAERTGT--KIEADIKAQDIYVYWSRGEKEK-VLNYARQ 327
QY 354 DVRYATYGAELKLPFAIQISVYTGLPDQVGAMSVGFLEWYLIRAAFKMKELVPRVER 413
DB 328 DAINTYLIAKELPHMYELSKMIRLPDVTVMGRGKQVDWMLLSEAKKI GELIAPNPPEH 387
QY 414 PEETVYGAIVLPLSGVHENIAVLDPSSNYPMIMIKYVGPDTLVRPGEKCGECGCNEAP 473
DB 388 -AESYEGAVLEPEPGLHENVACLDFAFMPSIMIAFNISPDY-----GCED-DCYEAP 440
QY 474 EVKHFRRPCPGFFKTIVLERLELRKRRVAENKKYFPDPSPEYRLLDERQKALKVLANASY 533
DB 441 EVGHRFKSPDFPFKLEMLLEKRELKVELKNLSPESEYKLLDIKQTKVLNTNGFY 500
QY 534 GYMGSGARWYCECAKAVTANGRHILIRTAIRAKLGLKVIYDGTDSLFTV---YDPEK 590
DB 501 GYMGWNLARWYCHPCAETANGRHIFIRTSAKIAESMGFKVLYGDTDSIFVTKAGMTKED 560
QY 591 VENFIKIIKEELGFEIKLEKVIYKRLFFTEAKRYAGLLEDGRIDIVGFEAVRGDWCELAK 650

DB 561 VDLRIDKLHEELPIQIEVDYISAFVVE-KKRYAGLTEDGRVYVKGLEVRGDMCBLAK 619
QY 651 EVQTKVVEIVLKTSEVNAKAVEYVKIVKELEEGKVPTEKLVIMKTLTKRLEETTEAPHV 710
DB 620 KVQREVEIVILKEKNPEKALSLVKDVIILRIKRGKVSLEEVYIYKLTGKPSKYSMQAHV 679
QY 711 VAAKMLSGAYRVSPGDKIGYIVVKGGRISQRAWPYFMVKD-----PSQ 755
DB 680 KAALKAREWGLIYFPVSSKIGYIVVKGSGNIGDRAVPIDLIEDFDGLENLRKTKSGIETIKK 739
QY 756 IDVTYVVDHQTIIIPALRIILGFGITEKLLKASATGOKTLPFLA 799
DB 740 LDKDYIIDNQIIPSVLRILERFGYTEASLKGSS--QMSLDSFFS 781

Search completed: November 25, 2003, 14:58:26
Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2003, 15:00:00 ; Search time 37 Seconds
(without alignments)
4002.918 Million cell updates/sec

Title: US-10-034-849-2
Perfect score: 4188
Sequence: 1 MTEVFTVLDSSYEVVKGEP.....LKASATGQKTLFDLAKSK 803

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 673684 seqs, 18443283 residues

Total number of hits satisfying chosen parameters: 673684

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4188	100.0	803	9 US-09-391-340-4	Sequence 4, Appli
2	4188	100.0	803	10 US-09-948-369-4	Sequence 4, Appli
3	4188	100.0	803	10 US-09-948-369-14	Sequence 14, Appli
4	4188	100.0	803	14 US-10-034-849-2	Sequence 2, Appli
5	4188	100.0	803	14 US-10-034-849-2	Sequence 2, Appli
6	4188	100.0	803	14 US-10-029-382-2	Sequence 2, Appli
7	4164	99.4	803	10 US-09-948-369-16	Sequence 16, Appli
8	1563	37.3	788	9 US-09-391-340-6	Sequence 6, Appli
9	1563	37.3	788	10 US-09-948-369-6	Sequence 6, Appli
10	1249.5	29.8	773	12 US-10-227-110-37	Sequence 37, Appli
11	1249.5	29.8	773	12 US-10-208-508-37	Sequence 37, Appli
12	1248	29.8	775	12 US-10-227-110-67	Sequence 67, Appli
13	1248	29.8	775	12 US-10-208-508-67	Sequence 67, Appli
14	1246.5	29.8	775	15 US-10-345-205-6	Sequence 6, Appli
15	1245.5	29.7	773	12 US-10-227-110-29	Sequence 29, Appli

15	1245.5	29.7	773	12	US-10-208-508-29	Sequence 29, Appli
17	1244.5	29.7	773	12	US-10-227-110-38	Sequence 38, Appli
18	1244.5	29.7	773	12	US-10-208-508-38	Sequence 38, Appli
19	1244.5	29.7	775	15	US-10-345-205-7	Sequence 7, Appli
20	1244	29.7	775	12	US-10-227-110-59	Sequence 59, Appli
21	1244	29.7	775	12	US-10-227-110-113	Sequence 113, App
22	1244	29.7	775	12	US-10-208-508-59	Sequence 59, Appli
23	1244	29.7	775	12	US-10-208-508-113	Sequence 113, App
24	1243	29.7	775	12	US-10-227-110-68	Sequence 68, Appli
25	1243	29.7	775	12	US-10-208-508-68	Sequence 68, Appli
26	1242.5	29.7	774	9	US-09-852-922-2	Sequence 2, Appli
27	1242.5	29.7	774	12	US-10-126-757A-7	Sequence 7, Appli
28	1241.5	29.6	773	12	US-10-227-110-73	Sequence 73, Appli
29	1241.5	29.6	773	12	US-10-208-508-73	Sequence 73, Appli
30	1241.5	29.6	775	15	US-10-345-205-1	Sequence 1, Appli
31	1240.5	29.6	773	12	US-10-227-110-32	Sequence 32, Appli
32	1240.5	29.6	773	12	US-10-208-508-32	Sequence 32, Appli
33	1240.5	29.6	775	12	US-10-227-110-27	Sequence 27, Appli
34	1240.5	29.6	775	12	US-10-208-508-37	Sequence 37, Appli
35	1240	29.6	775	12	US-10-227-110-85	Sequence 85, Appli
36	1240	29.6	775	12	US-10-208-508-85	Sequence 85, Appli
37	1239.5	29.6	773	12	US-10-227-110-35	Sequence 35, Appli
38	1239.5	29.6	773	12	US-10-227-110-74	Sequence 74, Appli
39	1239.5	29.6	773	12	US-10-227-110-76	Sequence 76, Appli
40	1239.5	29.6	773	12	US-10-208-508-35	Sequence 35, Appli
41	1239.5	29.6	773	12	US-10-208-508-74	Sequence 74, Appli
42	1239.5	29.6	773	12	US-10-208-508-76	Sequence 76, Appli
43	1239	29.6	775	12	US-10-227-110-62	Sequence 62, Appli
44	1239	29.6	775	12	US-10-208-508-62	Sequence 62, Appli
45	1238.5	29.6	773	12	US-10-227-110-75	Sequence 75, Appli

ALIGNMENTS

RESULT 1

US-09-391-340-4
; Sequence 4, Application US/09391340A
; Patent NO. US20020013455A1
; GENERAL INFORMATION:
; APPLICANT: Callen, Walter
; APPLICANT: Mather, Eric
; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
; FILE REFERENCE: 09010/027001
; CURRENT APPLICATION NUMBER: US/09/391,340A
; CURRENT FILING DATE: 1999-09-07
; EARLIER APPLICATION NUMBER: US 08/907,166
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 803
; TYPE: PRT
; ORGANISM: Pyrolobus fumarius
US-09-391-340-4

Query Match	100.0%;	Score 4188;	DB 9;	Length 803;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches	803;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
			0;	
Qy	1	MTEVFTVLDSSYEVVKGEPQVIWGIAENGERVLDSSFRPFYVALLAPGADPKQVAQ	60	
Db	1	MTEVFTVLDSSYEVVKGEPQVIWGIAENGERVLDSSFRPFYVALLAPGADPKQVAQ	60	
Qy	61	RIRALSRPKSPIIGVEDDKKFKYGRPRRVLRTVLPFAVREYRELKKNVGVGVEDVLEAD	120	
Db	61	RIRALSRPKSPIIGVEDDKKFKYGRPRRVLRTVLPFAVREYRELKKNVGVGVEDVLEAD	120	
Qy	121	IRFAMRYLIDHDLFPPTWYRVEABPLENKGFRVVKYLVKSRPELYGEALPTKLPDL	180	
Db	121	IRFAMRYLIDHDLFPPTWYRVEABPLENKGFRVVKYLVKSRPELYGEALPTKLPDL	180	
Qy	181	RIAFDIEVYKSGSRPRDPVIVIAVKTDGDEVLFIAEGKDKRPIREFVEYVKRYD	240	

```
Db 181 RILAFDIEVSKQSPRPDPVIVIAVKTDDGDEVLFIAEGKDRKPIREFVEVVKRYD 240
QY 241 PDIIYGVNNHFDWYLLRRARILGKLDVTRRVGAEPPTS VHGHVSVPGRNLVDLYDYA 300
Db 241 PDIIYGVNNHFDWYLLRRARILGKLDVTRRVGAEPPTS VHGHVSVPGRNLVDLYDYA 300
QY 301 EWPRIKIKSLSEVAEYLGVMKKSBRVLIINWWEI PDYDDPKRPLILQYARDVRYTG 360
Db 301 EWPRIKIKSLSEVAEYLGVMKKSBRVLIINWWEI PDYDDPKRPLILQYARDVRYTG 360
QY 361 LAEKILPFAIQLSYVTGLPDQVGAMSVGRLEWLIIRAAAFOMKELVNNRVERPSETYRG 420
Db 361 LAEKILPFAIQLSYVTGLPDQVGAMSVGRLEWLIIRAAAFOMKELVNNRVERPSETYRG 420
QY 421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYNGVDPDTLVRPGKCGCGCWEAPEVKHRFR 480
Db 421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYNGVDPDTLVRPGKCGCGCWEAPEVKHRFR 480
QY 481 RCPGPFKTVLERLLELRKRVRAEMKKYPPDSPERYLLDEROKALKVLANASYGMWSG 540
Db 481 RCPGPFKTVLERLLELRKRVRAEMKKYPPDSPERYLLDEROKALKVLANASYGMWSG 540
QY 541 ARWYCRCAKAVTANGSHLIRTAINTARKLGLKVIYGDTSLSFVYDPEKVENFIKIKE 600
Db 541 ARWYCRCAKAVTANGSHLIRTAINTARKLGLKVIYGDTSLSFVYDPEKVENFIKIKE 600
QY 601 ELGFEIKLEKVKRYLFFTEAKRYAGLLEDGRIDIVGFEAVRGDMCELAKEVQTKVVEIV 660
Db 601 ELGFEIKLEKVKRYLFFTEAKRYAGLLEDGRIDIVGFEAVRGDMCELAKEVQTKVVEIV 660
QY 661 LKTSVNKAVEYVKRIVKELEEGKVPTEKLVWTKLSKRELEYYTEAPHVVAAKRMLSAG 720
Db 661 LKTSVNKAVEYVKRIVKELEEGKVPTEKLVWTKLSKRELEYYTEAPHVVAAKRMLSAG 720
QY 721 YRVSPGDKIGYIVKGGGRISQRAWPFYMWKDPQSIDVTYVYDHOIIPAAALRILGYGIT 780
Db 721 YRVSPGDKIGYIVKGGGRISQRAWPFYMWKDPQSIDVTYVYDHOIIPAAALRILGYGIT 780
QY 781 EKKLKASATGOKTLDFLAKKSK 803
Db 781 EKKLKASATGOKTLDFLAKKSK 803
```

RESULT 2

US-09-948-369-4

; Sequence 4, Application US/09948369

; Patent No. US20020132243A1

; GENERAL INFORMATION:

; APPLICANT: DIVERSA CORPORATION

; APPLICANT: CALLEN, Walter

; APPLICANT: MATHUR, Eric

; TITLE OF INVENTION: ENZYMES HAVING HIGH TEMPERATURE POLYMERASE ACTIVITY AND METHODS

; FILE REFERENCE: DIVER1350-3

; CURRENT FILING DATE: 2001-09-06

; PRIOR FILING DATE: 2000-09-06

; PRIOR FILING DATE: 1999-09-07

; PRIOR FILING DATE: 1997-08-06

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 4

; LENGTH: 803

; TYPE: PRT

; ORGANISM: Pyrollobus fumarius

US-09-948-369-4

Query Match

Best Local Similarity 100.0%; Score 4188; DB 10; Length 803;

; Pred. No. 0;

```
Matches 803; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTEVVFTVLSSSEVVEVKGSPQVIINGIAENGERVWLIIDRSFRPYFYALLAPADPKQVAQ 60
Db 1 MTEVVFTVLSSSEVVEVKGSPQVIINGIAENGERVWLIIDRSFRPYFYALLAPADPKQVAQ 60
QY 61 RIRALSPPKSPILIGVEDDKKYPGRPRVLRIRTVLPEAVREYRELKNNVDGVEDVLEAD 120
Db 61 RIRALSPPKSPILIGVEDDKKYPGRPRVLRIRTVLPEAVREYRELKNNVDGVEDVLEAD 120
QY 121 IRPAMRYLIDHDLFPFTWYRVEAPELENKMGFRVDKYLVLKSRPEPIYGEALAPTLPDL 180
Db 121 IRPAMRYLIDHDLFPFTWYRVEAPELENKMGFRVDKYLVLKSRPEPIYGEALAPTLPDL 180
QY 181 RILAFDIEVSKQSPRPDPVIVIAVKTDDGDEVLFIAEGKDRKPIREFVEVVKRYD 240
Db 181 RILAFDIEVSKQSPRPDPVIVIAVKTDDGDEVLFIAEGKDRKPIREFVEVVKRYD 240
QY 241 PDIIYGVNNHFDWYLLRRARILGKLDVTRRVGAEPPTS VHGHVSVPGRNLVDLYDYA 300
Db 241 PDIIYGVNNHFDWYLLRRARILGKLDVTRRVGAEPPTS VHGHVSVPGRNLVDLYDYA 300
QY 301 EWPRIKIKSLSEVAEYLGVMKKSBRVLIINWWEI PDYDDPKRPLILQYARDVRYTG 360
Db 301 EWPRIKIKSLSEVAEYLGVMKKSBRVLIINWWEI PDYDDPKRPLILQYARDVRYTG 360
QY 361 LAEKILPFAIQLSYVTGLPDQVGAMSVGRLEWLIIRAAAFOMKELVNNRVERPSETYRG 420
Db 361 LAEKILPFAIQLSYVTGLPDQVGAMSVGRLEWLIIRAAAFOMKELVNNRVERPSETYRG 420
QY 421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYNGVDPDTLVRPGKCGCGCWEAPEVKHRFR 480
Db 421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYNGVDPDTLVRPGKCGCGCWEAPEVKHRFR 480
QY 481 RCPGPFKTVLERLLELRKRVRAEMKKYPPDSPERYLLDEROKALKVLANASYGMWSG 540
Db 481 RCPGPFKTVLERLLELRKRVRAEMKKYPPDSPERYLLDEROKALKVLANASYGMWSG 540
QY 541 ARWYCRCAKAVTANGSHLIRTAINTARKLGLKVIYGDTSLSFVYDPEKVENFIKIKE 600
Db 541 ARWYCRCAKAVTANGSHLIRTAINTARKLGLKVIYGDTSLSFVYDPEKVENFIKIKE 600
QY 601 ELGFEIKLEKVKRYLFFTEAKRYAGLLEDGRIDIVGFEAVRGDMCELAKEVQTKVVEIV 660
Db 601 ELGFEIKLEKVKRYLFFTEAKRYAGLLEDGRIDIVGFEAVRGDMCELAKEVQTKVVEIV 660
QY 661 LKTSVNKAVEYVKRIVKELEEGKVPTEKLVWTKLSKRELEYYTEAPHVVAAKRMLSAG 720
Db 661 LKTSVNKAVEYVKRIVKELEEGKVPTEKLVWTKLSKRELEYYTEAPHVVAAKRMLSAG 720
QY 721 YRVSPGDKIGYIVKGGGRISQRAWPFYMWKDPQSIDVTYVYDHOIIPAAALRILGYGIT 780
Db 721 YRVSPGDKIGYIVKGGGRISQRAWPFYMWKDPQSIDVTYVYDHOIIPAAALRILGYGIT 780
QY 781 EKKLKASATGOKTLDFLAKKSK 803
Db 781 EKKLKASATGOKTLDFLAKKSK 803
```

RESULT 3

US-09-948-369-14

; Sequence 14, Application US/09948369

; Patent No. US20020132243A1

; GENERAL INFORMATION:

; APPLICANT: DIVERSA CORPORATION

; APPLICANT: CALLEN, Walter

; APPLICANT: MATHUR, Eric

; TITLE OF INVENTION: ENZYMES HAVING HIGH TEMPERATURE POLYMERASE ACTIVITY AND METHODS

; FILE REFERENCE: DIVER1350-3

; CURRENT FILING DATE: 2001-09-06

; PRIOR FILING DATE: 2000-09-06

; PRIOR FILING DATE: 1997-08-06

;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: US 09/391,340
;; PRIOR FILING DATE: 1999-09-07
;; PRIOR APPLICATION NUMBER: US 08/907,166
;; PRIOR FILING DATE: 1997-08-06
;; NUMBER OF SEQ ID NOS: 16
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 14
;; LENGTH: 803
;; TYPE: PRT
;; ORGANISM: Pyrolobus fumarius
US-09-948-369-14

Query Match 100.0%; Score 4188; DB 10; Length 803;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 803; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEVFTVLDSSYEVVGKEPQVIINGIAENGERVVLIDRSFRPYFYALLAPGADPKQVAQ 60
DB 1 MTEVFTVLDSSYEVVGKEPQVIINGIAENGERVVLIDRSFRPYFYALLAPGADPKQVAQ 60

QY 61 RIRALSRPKSPIIGVEDDKRKYFGRPRVLRIRTVLPEAVRYRELKVNVDGVEDVLEAD 120
DB 61 RIRALSRPKSPIIGVEDDKRKYFGRPRVLRIRTVLPEAVRYRELKVNVDGVEDVLEAD 120

QY 121 IRPAMRYLIDHDLFPFTWYVVEAEPLKNGFRVDKYLKSRPEPLYGEALAPTLPDL 180
DB 121 IRPAMRYLIDHDLFPFTWYVVEAEPLKNGFRVDKYLKSRPEPLYGEALAPTLPDL 180

QY 181 RILAFDIEVYSKQSPRPDPVIVIAVKTDDGDEVLFIAEGKDDRKPIREVEVVKRYD 240
DB 181 RILAFDIEVYSKQSPRPDPVIVIAVKTDDGDEVLFIAEGKDDRKPIREVEVVKRYD 240

QY 241 PDIIIVGNNHFDWPLYLLRRARILGIKLDVTRRVGAEPPTSVGHVSVSGRLNVDLYDA 300
DB 241 PDIIIVGNNHFDWPLYLLRRARILGIKLDVTRRVGAEPPTSVGHVSVSGRLNVDLYDA 300

QY 301 EEMPEIKTKSLSEVAEVLGVNKKSERVIINWWEIPDYDDPKKPLILQYARDDVRATYG 360
DB 301 EEMPEIKTKSLSEVAEVLGVNKKSERVIINWWEIPDYDDPKKPLILQYARDDVRATYG 360

QY 361 LAEKILPFAIQLSYVTGLPDQVGAMSVGRLEWYLIRAAFKMKELVFNVRPERPEYRG 420
DB 361 LAEKILPFAIQLSYVTGLPDQVGAMSVGRLEWYLIRAAFKMKELVFNVRPERPEYRG 420

QY 421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYNGVGPDTLVRPGEKCGCGCWEAPEVKHRFR 480
DB 421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYNGVGPDTLVRPGEKCGCGCWEAPEVKHRFR 480

QY 481 RCPGPFKTVLERLELRKRVRAEMKKYPPDSPYELLDEROKALKVLANASYGMWSG 540
DB 481 RCPGPFKTVLERLELRKRVRAEMKKYPPDSPYELLDEROKALKVLANASYGMWSG 540

QY 541 ARMYCRECAKAVTANGRHILRTAINIARKLGLKVIYGDTSDFVYDPEKVENFIKIKE 600
DB 541 ARMYCRECAKAVTANGRHILRTAINIARKLGLKVIYGDTSDFVYDPEKVENFIKIKE 600

QY 601 ELGFEIKLEKVKYKRLFFTEAKRYAGLLEDGRIDIVGFEAVRGDWCBLAKEVQTKVEIV 660
DB 601 ELGFEIKLEKVKYKRLFFTEAKRYAGLLEDGRIDIVGFEAVRGDWCBLAKEVQTKVEIV 660

QY 661 LKTSSEVNKAVEYVRKIVKELEGKVPFIEKLVTWKTLKLELVEYTTTEAPHVAAKRLSAG 720
DB 661 LKTSSEVNKAVEYVRKIVKELEGKVPFIEKLVTWKTLKLELVEYTTTEAPHVAAKRLSAG 720

QY 721 YRVSPODKIGYIVKGGGRISQRAWPYFMVQPSQIDVTYYVDHQIIPALAILGFGIT 780
DB 721 YRVSPODKIGYIVKGGGRISQRAWPYFMVQPSQIDVTYYVDHQIIPALAILGFGIT 780

QY 781 EKKLKASATGQKTLFDPLAKSK 803
DB 781 EKKLKASATGQKTLFDPLAKSK 803

RESULT 4

US-10-034-849-2
;; Sequence 2, Application US/10034849
;; Publication No. US20020115108A1
;; GENERAL INFORMATION:
;; APPLICANT: DIVERSA CORPORATION
;; APPLICANT: CALLEN, Walter
;; APPLICANT: MATHUR, Eric
;; TITLE OF INVENTION: ENZYMES HAVING HIGH TEMPERATURE POLYMERASE ACTIVITY
;; TITLE OF INVENTION: AND METHODS OF USE
;; TITLE OF INVENTION: THEREOF
;; FILE REFERENCE: DIVER1350-2
;; CURRENT APPLICATION NUMBER: US/10/034,849
;; CURRENT FILING DATE: 2001-12-21
;; PRIOR APPLICATION NUMBER: US/09/656,309
;; PRIOR FILING DATE: 2000-09-06
;; NUMBER OF SEQ ID NOS: 2
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 2
;; LENGTH: 803
;; TYPE: PRT
;; ORGANISM: Pyrolobus fumarius
US-10-034-849-2

Query Match 100.0%; Score 4188; DB 14; Length 803;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 803; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEVFTVLDSSYEVVGKEPQVIINGIAENGERVVLIDRSFRPYFYALLAPGADPKQVAQ 60
DB 1 MTEVFTVLDSSYEVVGKEPQVIINGIAENGERVVLIDRSFRPYFYALLAPGADPKQVAQ 60

QY 61 RIRALSRPKSPIIGVEDDKRKYFGRPRVLRIRTVLPEAVRYRELKVNVDGVEDVLEAD 120
DB 61 RIRALSRPKSPIIGVEDDKRKYFGRPRVLRIRTVLPEAVRYRELKVNVDGVEDVLEAD 120

QY 121 IRPAMRYLIDHDLFPFTWYVVEAEPLKNGFRVDKYLKSRPEPLYGEALAPTLPDL 180
DB 121 IRPAMRYLIDHDLFPFTWYVVEAEPLKNGFRVDKYLKSRPEPLYGEALAPTLPDL 180

QY 181 RILAFDIEVYSKQSPRPDPVIVIAVKTDDGDEVLFIAEGKDDRKPIREVEVVKRYD 240
DB 181 RILAFDIEVYSKQSPRPDPVIVIAVKTDDGDEVLFIAEGKDDRKPIREVEVVKRYD 240

QY 241 PDIIIVGNNHFDWPLYLLRRARILGIKLDVTRRVGAEPPTSVGHVSVSGRLNVDLYDA 300
DB 241 PDIIIVGNNHFDWPLYLLRRARILGIKLDVTRRVGAEPPTSVGHVSVSGRLNVDLYDA 300

QY 301 EEMPEIKTKSLSEVAEVLGVNKKSERVIINWWEIPDYDDPKKPLILQYARDDVRATYG 360
DB 301 EEMPEIKTKSLSEVAEVLGVNKKSERVIINWWEIPDYDDPKKPLILQYARDDVRATYG 360

QY 361 LAEKILPFAIQLSYVTGLPDQVGAMSVGRLEWYLIRAAFKMKELVFNVRPERPEYRG 420
DB 361 LAEKILPFAIQLSYVTGLPDQVGAMSVGRLEWYLIRAAFKMKELVFNVRPERPEYRG 420

QY 421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYNGVGPDTLVRPGEKCGCGCWEAPEVKHRFR 480
DB 421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYNGVGPDTLVRPGEKCGCGCWEAPEVKHRFR 480

QY 481 RCPGPFKTVLERLELRKRVRAEMKKYPPDSPYELLDEROKALKVLANASYGMWSG 540
DB 481 RCPGPFKTVLERLELRKRVRAEMKKYPPDSPYELLDEROKALKVLANASYGMWSG 540

QY 541 ARMYCRECAKAVTANGRHILRTAINIARKLGLKVIYGDTSDFVYDPEKVENFIKIKE 600
DB 541 ARMYCRECAKAVTANGRHILRTAINIARKLGLKVIYGDTSDFVYDPEKVENFIKIKE 600

QY 601 ELGFEIKLEKVKYKRLFFTEAKRYAGLLEDGRIDIVGFEAVRGDWCBLAKEVQTKVEIV 660
DB 601 ELGFEIKLEKVKYKRLFFTEAKRYAGLLEDGRIDIVGFEAVRGDWCBLAKEVQTKVEIV 660

```
QY 661 LKTSVNKAVEYVRKIVKSELSEGVPIEKLVWTKLSKRLSEYTTTEAPHVVAARMLSAG 720
DB 661 LKTSVNKAVEYVRKIVKSELSEGVPIEKLVWTKLSKRLSEYTTTEAPHVVAARMLSAG 720
QY 721 YRVSPGDKIGYVIVKGGGRISQRAWPYFMVKDPSQIDVTYVVDHQIIPAAALRILGYFGIT 780
DB 721 YRVSPGDKIGYVIVKGGGRISQRAWPYFMVKDPSQIDVTYVVDHQIIPAAALRILGYFGIT 780
QY 781 EKKLKASATGQKTLDFDLAKSK 803
DB 781 EKKLKASATGQKTLDFDLAKSK 803

RESULT 5
US-10-034-621-2
; Sequence 2, Application US/10034621
; Publication No. US20020132997A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: CALLEN, Walter
; APPLICANT: MATHUR, Eric
; TITLE OF INVENTION: ENZYMES HAVING HIGH TEMPERATURE POLYMERASE ACTIVITY
; TITLE OF INVENTION: AND METHODS OF USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: DIVER1350-2
; CURRENT APPLICATION NUMBER: US/10/034,621
; CURRENT FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US/09/656,309
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 803
; TYPE: PRT
; ORGANISM: Pyrolobus fumarius
US-10-034-621-2

Query Match 100.0%; Score 4188; DB 14; Length 803;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 803; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEVFTVLDSSEYVGVKPEQVIINGIAENGERVVLIDRSFRPYFYALLAPGADPKQVAQ 60
DB 1 MTEVFTVLDSSEYVGVKPEQVIINGIAENGERVVLIDRSFRPYFYALLAPGADPKQVAQ 60
QY 61 RIRALSRLSPKSPIIGVEDDKKRYFGRPRVLRIRTVLPEAVREYRELKVNVDGVEDVLEAD 120
DB 61 RIRALSRLSPKSPIIGVEDDKKRYFGRPRVLRIRTVLPEAVREYRELKVNVDGVEDVLEAD 120
QY 121 IRFAMRYLIDHDLFPFTWYRVAEAPLENKMGFRVDKVLVKSRLPELYGEALAPTKLPDL 180
DB 121 IRFAMRYLIDHDLFPFTWYRVAEAPLENKMGFRVDKVLVKSRLPELYGEALAPTKLPDL 180
QY 181 RILAEPIEYVSKQSGPRPRDPVIVIAVKTDDGDEVLFIAEGKDDRKPIREFVEYVKRYD 240
DB 181 RILAEPIEYVSKQSGPRPRDPVIVIAVKTDDGDEVLFIAEGKDDRKPIREFVEYVKRYD 240
QY 241 PDIIIVGNNHFDMPYLLRRARILGKLDVTRVGAEPPTS VHGVSVPGRLNVDLYDA 300
DB 241 PDIIIVGNNHFDMPYLLRRARILGKLDVTRVGAEPPTS VHGVSVPGRLNVDLYDA 300
QY 301 EEMPEIKISLEVAEYLVGMKSERVIINWWEIPDYWDPPKRPILLIQYARDVVRATYG 360
DB 301 EEMPEIKISLEVAEYLVGMKSERVIINWWEIPDYWDPPKRPILLIQYARDVVRATYG 360
QY 361 LAEKILFFAIQLSVYVGLDQVGMASVGRLEWILIRAAFKMKELVNPVRPERSETYRG 420
DB 361 LAEKILFFAIQLSVYVGLDQVGMASVGRLEWILIRAAFKMKELVNPVRPERSETYRG 420
QY 421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYNGVPTLVRPBGKCGCGCWEAPEVKHRFR 480
DB 421 AIVLEPLRGVHENIAVLDFSSMYPNIMIKYNGVPTLVRPBGKCGCGCWEAPEVKHRFR 480
```

```
QY 481 RCPGPFKTVLERLLRLRRVRAEMKKYPPDPSPEYRLDROKALKVLANASYGWMGSG 540
DB 481 RCPGPFKTVLERLLRLRRVRAEMKKYPPDPSPEYRLDROKALKVLANASYGWMGSG 540
QY 541 ARWYCECAKAVTAWGRHLIRTAINTARKLGLKVIYVGTDSLFVTVYDPEKVENFIKIKE 600
DB 541 ARWYCECAKAVTAWGRHLIRTAINTARKLGLKVIYVGTDSLFVTVYDPEKVENFIKIKE 600
QY 601 ELGFBIEKLEKYVKKLFFTEAKRYAGLLEDGRIDIVGFEAVRGDNCCLAKEVQTKVEIV 660
DB 601 ELGFBIEKLEKYVKKLFFTEAKRYAGLLEDGRIDIVGFEAVRGDNCCLAKEVQTKVEIV 660
QY 661 LKTSVNKAVEYVRKIVKSELSEGVPIEKLVWTKLSKRLSEYTTTEAPHVVAARMLSAG 720
DB 661 LKTSVNKAVEYVRKIVKSELSEGVPIEKLVWTKLSKRLSEYTTTEAPHVVAARMLSAG 720
QY 721 YRVSPGDKIGYVIVKGGGRISQRAWPYFMVKDPSQIDVTYVVDHQIIPAAALRILGYFGIT 780
DB 721 YRVSPGDKIGYVIVKGGGRISQRAWPYFMVKDPSQIDVTYVVDHQIIPAAALRILGYFGIT 780
QY 781 EKKLKASATGQKTLDFDLAKSK 803
DB 781 EKKLKASATGQKTLDFDLAKSK 803

RESULT 6
US-10-029-382-2
; Sequence 2, Application US/10029382
; Publication No. US20020164618A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: CALLEN, Walter
; APPLICANT: MATHUR, Eric
; TITLE OF INVENTION: ENZYMES HAVING HIGH TEMPERATURE POLYMERASE ACTIVITY AND METHODS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: DIVER1350-2
; CURRENT APPLICATION NUMBER: US/10/029,382
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US/09/656,309
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 803
; TYPE: PRT
; ORGANISM: Pyrolobus fumarius
US-10-029-382-2

Query Match 100.0%; Score 4188; DB 14; Length 803;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 803; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTEVFTVLDSSEYVGVKPEQVIINGIAENGERVVLIDRSFRPYFYALLAPGADPKQVAQ 60
DB 1 MTEVFTVLDSSEYVGVKPEQVIINGIAENGERVVLIDRSFRPYFYALLAPGADPKQVAQ 60
QY 61 RIRALSRLSPKSPIIGVEDDKKRYFGRPRVLRIRTVLPEAVREYRELKVNVDGVEDVLEAD 120
DB 61 RIRALSRLSPKSPIIGVEDDKKRYFGRPRVLRIRTVLPEAVREYRELKVNVDGVEDVLEAD 120
QY 121 IRFAMRYLIDHDLFPFTWYRVAEAPLENKMGFRVDKVLVKSRLPELYGEALAPTKLPDL 180
DB 121 IRFAMRYLIDHDLFPFTWYRVAEAPLENKMGFRVDKVLVKSRLPELYGEALAPTKLPDL 180
QY 181 RILAEPIEYVSKQSGPRPRDPVIVIAVKTDDGDEVLFIAEGKDDRKPIREFVEYVKRYD 240
DB 181 RILAEPIEYVSKQSGPRPRDPVIVIAVKTDDGDEVLFIAEGKDDRKPIREFVEYVKRYD 240
QY 241 PDIIIVGNNHFDMPYLLRRARILGKLDVTRVGAEPPTS VHGVSVPGRLNVDLYDA 300
DB 241 PDIIIVGNNHFDMPYLLRRARILGKLDVTRVGAEPPTS VHGVSVPGRLNVDLYDA 300
QY 301 EEMPEIKISLEVAEYLVGMKSERVIINWWEIPDYWDPPKRPILLIQYARDVVRATYG 360
```

Db 301 EEMPEIKIKSLEAEVAYLVGKMSKSERVINWWEIPDYMDPKRPLLLQYARDVYRATG 360
QY 361 LAEKILPFAIQLSYVTGLPLDQVGAMSVGRLEWYLIRAAFPKOKBLVNVNRPBETVRG 420
Db 361 LAEKILPFAIQLSYVTGLPLDQVGAMSVGRLEWYLIRAAFPKOKBLVNVNRPBETVRG 420
QY 421 AIVLEPLRGVHENIAVLDFSSMYNIMIKYNGVDPDTPVRPGEKCGCGWEAPEVKHRFR 480
Db 421 AIVLEPLRGVHENIAVLDFSSMYNIMIKYNGVDPDTPVRPGEKCGCGWEAPEVKHRFR 480
QY 481 RCPGPFKTVLERLLELRKRVRAEMKKYPPDPSPEYLLDERQKALKVLANASYGYMGWSG 540
Db 481 RCPGPFKTVLERLLELRKRVRAEMKKYPPDPSPEYLLDERQKALKVLANASYGYMGWSG 540
QY 541 ARWYCRECAKAVTANGRHILRTAINTARIKLGKLVYIGDTSLSFTVYDPEKVENFIKIIE 600
Db 541 ARWYCRECAKAVTANGRHILRTAINTARIKLGKLVYIGDTSLSFTVYDPEKVENFIKIIE 600
QY 601 ELGFEIKLEKVKYKRLFFTEAKRYAGLLEDDGRIDIVGFEAVRGDWCCELAKVOTKVEIV 660
Db 601 ELGFEIKLEKVKYKRLFFTEAKRYAGLLEDDGRIDIVGFEAVRGDWCCELAKVOTKVEIV 660
QY 661 LKTSSEVNAKAVEYVRKIVKELEBEGKVPKIEKLVWTKLSKLELEYTTEAPHVAAKRMLSAG 720
Db 661 LKTSSEVNAKAVEYVRKIVKELEBEGKVPKIEKLVWTKLSKLELEYTTEAPHVAAKRMLSAG 720
QY 721 YRVSPGDKIGYVIVKGGGRISQRAWPFYFVWVMDPSQIDVTYYVDHQQIIPALRIILGYFGIT 780
Db 721 YRVSPGDKIGYVIVKGGGRISQRAWPFYFVWVMDPSQIDVTYYVDHQQIIPALRIILGYFGIT 780
QY 781 EKKLKASATGQKTLDFDLAKKSK 803
Db 781 EKKLKASATGQKTLDFDLAKKSK 803

RESULT 7

US-09-948-369-16
; Sequence 16, Application US/09948369
; Patent No. US20020132243A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: CALLEN, Walter
; APPLICANT: MATHUR, Eric
; TITLE OF INVENTION: ENZYMES HAVING HIGH TEMPERATURE POLYMERASE ACTIVITY AND METHODS THEREOF
; FILE REFERENCE: DIVER1350-3
; CURRENT APPLICATION NUMBER: US/09/948,369
; PRIOR FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: US 09/656,309
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/391,340
; PRIOR FILING DATE: 1999-09-07
; PRIOR APPLICATION NUMBER: US 08/907,166
; PRIOR FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 16
; LENGTH: 803
; TYPE: PRT
; ORGANISM: Pyrolabus fumarius
US-09-948-369-16

Query Match 99.4%; Score 4164; DB 10; Length 803;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 797; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTEWFTVLDSSVEYVKGEPQVLIWGIAENGRVVLIDRSFRPYFVALLAPCADPKQVAQ 60
Db 1 MTEWFTVLDSSVEYVKGEPQVLIWGIAENGRVVLIDRSFRPYFVALLAPCADPKQVAQ 60
QY 61 RIRALSRPKSPIGVEDDKRKYFCRPRVLIRITVLPEAVREYRELKVNVDGVEDVLEAD 120

Db 61 RIRALSRPKSPIGVEDDKRKYFCRPRVLIRITVLPEAVREYRELKVNVDGVEDVLEAD 120
QY 121 IRFAMRYLIDHDLFPFTWYRVEAPLENKMGFRVDKVLVKSRRPEPLYGALAPTKLPDL 180
Db 121 IRFAMRYLIDHDLFPFTWYRVEAPLENKMGFRVDKVLVKSRRPEPLYGALAPTKLPDL 180
QY 181 RILAPDLEIVSKQSGPRPERDPIVIAVKTDDGDEVLFIAEGKDDRPPIREFVEYVKRYD 240
Db 181 RILAPDLEIVSKQSGPRPERDPIVIAVKTDDGDEVLFIAEGKDDRPPIREFVEYVKRYD 240
QY 241 PDIIIVGNNHFDWPLYLLRRARILGILKIDVTRRVGASPTTSVHGSHVSPGRLNVDLYDA 300
Db 241 PDIIIVGNNHFDWPLYLLRRARILGILKIDVTRRVGASPTTSVHGSHVSPGRLNVDLYDA 300
QY 301 EEMPEIKIKSLEAEVAYLVGKMSKSERVINWWEIPDYMDPKRPLLLQYARDVYRATG 360
Db 301 EEMPEIKIKSLEAEVAYLVGKMSKSERVINWWEIPDYMDPKRPLLLQYARDVYRATG 360
QY 361 LAEKILPFAIQLSYVTGLPLDQVGAMSVGRLEWYLIRAAFPKOKBLVNVNRPBETVRG 420
Db 361 LAEKILPFAIQLSYVTGLPLDQVGAMSVGRLEWYLIRAAFPKOKBLVNVNRPBETVRG 420
QY 421 AIVLEPLRGVHENIAVLDFSSMYNIMIKYNGVDPDTPVRPGEKCGCGWEAPEVKHRFR 480
Db 421 AIVLEPLRGVHENIAVLDFSSMYNIMIKYNGVDPDTPVRPGEKCGCGWEAPEVKHRFR 480
QY 481 RCPGPFKTVLERLLELRKRVRAEMKKYPPDPSPEYLLDERQKALKVLANASYGYMGWSG 540
Db 481 RCPGPFKTVLERLLELRKRVRAEMKKYPPDPSPEYLLDERQKALKVLANASYGYMGWSG 540
QY 541 ARWYCRECAKAVTANGRHILRTAINTARIKLGKLVYIGDTSLSFTVYDPEKVENFIKIIE 600
Db 541 ARWYCRECAKAVTANGRHILRTAINTARIKLGKLVYIGDTSLSFTVYDPEKVENFIKIIE 600
QY 601 ELGFEIKLEKVKYKRLFFTEAKRYAGLLEDDGRIDIVGFEAVRGDWCCELAKVOTKVEIV 660
Db 601 ELGFEIKLEKVKYKRLFFTEAKRYAGLLEDDGRIDIVGFEAVRGDWCCELAKVOTKVEIV 660
QY 661 LKTSSEVNAKAVEYVRKIVKELEBEGKVPKIEKLVWTKLSKLELEYTTEAPHVAAKRMLSAG 720
Db 661 LKTSSEVNAKAVEYVRKIVKELEBEGKVPKIEKLVWTKLSKLELEYTTEAPHVAAKRMLSAG 720
QY 721 YRVSPGDKIGYVIVKGGGRISQRAWPFYFVWVMDPSQIDVTYYVDHQQIIPALRIILGYFGIT 780
Db 721 YRVSPGDKIGYVIVKGGGRISQRAWPFYFVWVMDPSQIDVTYYVDHQQIIPALRIILGYFGIT 780
QY 781 EKKLKASATGQKTLDFDLAKKSK 803
Db 781 EKKLKASATGQKTLDFDLAKKSK 803

RESULT 8

US-09-391-340-6
; Sequence 6, Application US/09391340A
; Patent No. US20020013455A1
; GENERAL INFORMATION:
; APPLICANT: Callen, Walter
; APPLICANT: Mather, Eric
; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
; FILE REFERENCE: 09010/027001
; CURRENT APPLICATION NUMBER: US/09/391,340A
; CURRENT FILING DATE: 1999-09-07
; EARLIER APPLICATION NUMBER: US 08/907,166
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 788
; TYPE: PRT
; ORGANISM: Archaeoglobus lithotrophicus
US-09-391-340-6

Query Match 37.3%; Score 1563; DB 9; Length 788;

Db 427 GCEYDVAPQVGHAFCKDFGFTPSLGLDLLEERQVKKMA-TIDPIEKLLDYRQA 485
 QY 525 LKVLNAYSYGVMGSGARWYCRCAKAVTANGHLLRTAI-NIAKGLGKLVYGTDSL 583
 Db 486 IKILANSFYGYGAKARWYCKEABSVTANGROYTETIREIEKFGFKVLYADTGGF 545
 QY 594 VYDPEKVENFIKILKEELGF-----EIKLEKVKYKRLPTEAKKRYAGLLDGRID 634
 Db 546 ATIPGADAEVYKKAKEFLDYINAKLPGLLEVEGYKGFV-TKKYAVIDEEDKIT 604
 QY 635 IVGPEAVRGDCELAKEVQTKVETIVLKTSEVNAKVEYKRVKIVKELEBEGVPIELKVIN 694
 Db 605 TRGLEIVRRDSEIAKETQARVLSAILKHGDVEEVRIVKEVTEKLSKYVEPEKLVYE 664
 QY 695 TSKRLSEYTTTEAPHHVAAKMSAGVRSFGDKIGVIVKGGGRISORAWPYFVKDPS 754
 Db 665 QITRDLKDYKATGHVAVAKLAARGIKIRPGTVISYVLKSGRIGDRAIP-PDEFDA 723
 QY 755 --QIDVTVYVDHQLIPALRILGVFGITEKKLKASGTGQKTLFDFLAKKS 802
 Db 724 KHKYDAEYIENQVLPFAVERILRAFGRYKEDLRYQKTRQVGLGAMLKPT 773

RESULT 12

US-10-227-110-67
 ; Sequence 67, Application US/10227110
 ; Publication No. US20030143577A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Stratagene
 ; TITLE OF INVENTION: HIGH FIDELITY DNA POLYMERASE COMPOSITIONS AND USES THEREFOR
 ; FILE REFERENCE: 25436/2155C
 ; CURRENT APPLICATION NUMBER: US/10/227,110
 ; CURRENT FILING DATE: 2002-08-23
 ; PRIOR APPLICATION NUMBER: 10/079,241
 ; PRIOR FILING DATE: 2002-02-20
 ; PRIOR APPLICATION NUMBER: 10/208,508
 ; PRIOR FILING DATE: 2002-07-30
 ; PRIOR APPLICATION NUMBER: 10/035,091
 ; PRIOR FILING DATE: 2001-12-21
 ; NUMBER OF SEQ ID NOS: 128
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 67
 ; LENGTH: 775
 ; TYPE: PRT
 ; ORGANISM: Pyrococcus sp.
 US-10-227-110-67

Query Match 29.8%; Score 1248; DB 12; Length 775;
 Best Local Similarity 36.1%; Pred. No. 2.7e-106;
 Matches 301; Conservative 154; Mismatches 281; Indels 98; Gaps 22;
 QY 8 VLDSSYEVVGKEPQVINGIAENGERVVLIDRSRPPYFVYALLAPGADPKQV---AQRIR 63
 Db 2 ILDADYITDGDGKPIIRIFK-KENGEEFKVEYDRNRPPIYALLKDDSQIDEVRKITAERHG 60
 QY 64 ALSRPSKPIIGVEDDKRYKFGRRVLRIRTVLPEAVREYRELKVNQGVGVLEADIRF 123
 Db 61 KIVR-----IIDAERKVKFLGRPIEVWRLYFEHPQDPAIRDKIRHSAVILFEYDIPF 116
 QY 124 AMRYLIDHDLFPFTWYRVEABPLENKMGFRVDKVLVYKSRPEPLYGSEALAPTKLPOLRIL 183
 Db 117 AKRYLIDKGLI-----PMEGD-----BELKLL 138
 QY 184 AFDIEVYKSGSPRPERDPVIVIAKTTDDGDEVL-----FIAEGDKDKPIREFVEY 235
 Db 139 AFDIETLHEGE-EPAKGPIMISYADEEENKVTWKIDLPVYVVSSEKWKIFLKV 197
 QY 236 VKRYDPDITVGNHHPDWYLLRRLRILGILKLDVTRRVGAETTSVHG-----VSVPSGL 292
 Db 198 IREKDPDVIITVNGDSFDPYLVKRAEKLGIKLPGRD-GSEPKQRLGDMTAVEIKGRI 256
 QY 293 NVDLVDYABENPEIKLSBVAEVLGWMKSERVINWWEIPOYWDPKPKPLLLQYAR 352

Db 257 HFDLXHVIRRTNLPTTYLEAVYEA-EGRPKEKVAH-EIAEARNETCKGLERVAKYSM 313
 QY 353 DVVRATYGLAKILIPAIQISYVTGLDQVGMSSVGRLEWVILIRAAAPMELVPNRVE 412
 Db 314 EDKAVTYELGSEFFFMBAQLSRLVGLDVBVRSSTGNLVWYLLKAYERNELAPNRPD 373
 QY 413 RPE-----ETRYGAIVLEPLRGVHENIAVLDFSSMYPNIMIKVNVGPTLVRPGEKCOE 466
 Db 374 EREYERRLRESYAGAVYKEPEKGLMEGLVSLDFRSLYPSIIITHVSPDTLNR----- 427
 QY 467 CQCEW---APEVKHFRRCPPGPFKTVLELLELLEKRAEMKYPDPSPFVRLDERQK 523
 Db 428 -GCRDYVAPEVGHKFKOPFGFTFSLKLLDDEOBEIKRWKCA-SKDIETKMLDYROR 485
 QY 524 ALKVLNAYSYGVMGSGARWYCRCAKAVTANGHLLR-TAINIARKLGLKVYGTDSL 582
 Db 486 AIKILANSYGYGVAKARWYCKEABSVTANGREXIEFVRKELEBEGFKPKVLYIDTDL 545
 QY 593 FVTY---DPEKVE---NFIKILKEELG--FEIKLEKVKYKRLPTEAKKRYAGLLDGR 633
 Db 546 YATIPCAKPEEIKKALEFVDYINAKLPGLLEVEGYKGFV-TKKYALIDBEGKI 604
 QY 634 DIVGPEAVRGDCELAKEVQTKVETIVLKTSEVNAKVEYKRVKIVKELEBEGVPIELKVIN 693
 Db 605 ITRGLEIVRRDSEIAKETQAKVLEALKHGNEEAVKIVKSTELSKYEIPPEKLVY 664
 QY 694 KTLSEKLEBYTTEAPHHVAAKMSAGVRSFGDKIGVIVKGGGRISORAWPYFVKD- 752
 Db 665 EQITRPLSHYKAIGHVAVAKLAARGVVRPGWVIGVYVLRDGDPISKRA---ILAERF 721
 QY 753 --PQIDVTVYVDHQLIPALRILGVFGITEKKLKASATGQKTLFDFLAKSK 803
 Db 722 LERKHKYDAEYIENQVLPFAVLRILEAFGRYKEDLRYQKTRQVGLGAMLKPK 775

RESULT 13

US-10-208-508-67
 ; Sequence 67, Application US/10208508
 ; Publication No. US20030180741A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Stratagene
 ; TITLE OF INVENTION: HIGH FIDELITY DNA POLYMERASE COMPOSITIONS AND USES THEREFOR
 ; FILE REFERENCE: 25436/2155B
 ; CURRENT APPLICATION NUMBER: US/10/208,508
 ; CURRENT FILING DATE: 2003-02-07
 ; PRIOR APPLICATION NUMBER: 10/079,241
 ; PRIOR FILING DATE: 2002-02-20
 ; PRIOR APPLICATION NUMBER: 10/035,091
 ; PRIOR FILING DATE: 2001-12-21
 ; NUMBER OF SEQ ID NOS: 128
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 67
 ; LENGTH: 775
 ; TYPE: PRT
 ; ORGANISM: Pyrococcus sp.
 US-10-208-508-67

Query Match 29.8%; Score 1248; DB 12; Length 775;
 Best Local Similarity 36.1%; Pred. No. 2.7e-106;
 Matches 301; Conservative 154; Mismatches 281; Indels 98; Gaps 22;
 QY 8 VLDSSYEVVGKEPQVINGIAENGERVVLIDRSRPPYFVYALLAPGADPKQV---AQRIR 63
 Db 2 ILDADYITDGDGKPIIRIFK-KENGEEFKVEYDRNRPPIYALLKDDSQIDEVRKITAERHG 60
 QY 64 ALSRPSKPIIGVEDDKRYKFGRRVLRIRTVLPEAVREYRELKVNQGVGVLEADIRF 123
 Db 61 KIVR-----IIDAERKVKFLGRPIEVWRLYFEHPQDPAIRDKIRHSAVILFEYDIPF 116
 QY 124 AMRYLIDHDLFPFTWYRVEABPLENKMGFRVDKVLVYKSRPEPLYGSEALAPTKLPOLRIL 183
 Db 117 AKRYLIDKGLI-----PMEGD-----BELKLL 138

QY 184 AFDIEVSKQSPRPDPVIVIAVKTDDGDEVL-----FIAEGKDRKPIREFVEY 235
 Db 139 AFDIEVHEGE-EFAGKPIIMISYADSEAKVITWKIDLPFYEVVSSEREMIKRFLKV 197
 QY 236 VRRYDPDIIIVGNNHDFWFLRRARILGILKLDVTRVGAEPPTTSVHGH---VSVPGRL 292
 Db 198 IREKDPDVIITNGSDFDLPYLKGAELGILKPLGRD-GSEPKQRLGDMTAVEIKGRI 256
 QY 293 NVLDYDABEMPEIKISLEVAEYLGVMKSERVINWMEIPDYWDPPKRLDLOQYAR 352
 Db 257 HFDLYHVRITINLPTYLEAVYSAI-FGKPKKYYAH--EIAEAMTGTGLERVAKYSM 313
 QY 353 DVVRATYGLAEKILPFAIQLSVVTGLPLDQVGMVGFRLWYLIRAAFMKELVNRVE 412
 Db 314 EDAKVITYELGREFPMEAGLSLNGQPLMDVSRSTGNLVEMYLKAYERNELAKPKD 373
 QY 413 RPE-----ETVRGAILVLEPLRGVHENTAVLDPSMYPNIMIKNVGPDILVRPGEKCGE 466
 Db 374 EREYERRLRESYAGAYVKEPEKGLWGLVSLDPRSLYPSIIITHNVSPDTLNRE-----427
 QY 467 CGCWE---APEVKHFRFCPPGFFKTVLERLLELRKSVRAEMKKYPPDSPEYRLDDEKQ 523
 Db 428 -GCREVDVAPVGHGKCFKDFGFIPLSLKRLDDEKQEKRAKWA-SKDPLEKMWLDYRQ 485
 QY 524 ALKLANASYGVMGSGARWYCRCAKAVTANGRHILIR-TAINIARKLGLKVYGDPSL 582
 Db 486 AIKILANSYGYGYAKARWYCKEASVTANGREYIEFVRKELEERFGFKVLYIDTDL 545
 QY 583 FVTY---DPEKVE---NFIKIKKEELG--FEIKLEKYKRLFFTEAKKRYAGLEDPRI 633
 Db 546 YATIPGAKPEBIIKKALFEVDYINAKPLGLELEVEGYVKGFFV-TKKIYALDDEGKI 604
 QY 634 DIVGFAVRGDMCELAKVQTKVBEIVLKTSEVNKAVEYVAKIVKELEBEGKVPLEKLV 693
 Db 605 ITRGLEIVRRDWESEIAKSTQAKVAILKHGNVBEAVKIVKEVTEKLSKYEIPPEKLV 664
 QY 694 KTLKRLBYTTEAHPHVAAKMELSGVYRSPGDKIGVIVKGGRIISQRAWPFYVVKD- 752
 Db 665 EQITRPLHYKALGHVAVAKKLAARGVKVPGMVIGIIVLRGDPISKRA---ILAEF 721
 QY 753 ---PSQIDVTYVVDHQIIPAAIRILGYFGITEKKLKAASATQKTLDFLAKSK 803
 Db 722 DLKHKYDAEYIENQVLPVAVLRILEAFGRKEDLRWQKTKQTGLTAWLNKKK 775

RESULT 14

US-10-345-205-6
 ; Sequence 6, Application US/10345205
 ; Publication No. US20030105615A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DOI, Hitofumi
 ; APPLICANT: Hiraki, Hideaki
 ; APPLICANT: Kanai, Akio
 ; TITLE OF INVENTION: A METHOD AND A SYSTEM FOR PREDICTING PROTEIN FUNCTIONAL
 ; TITLE OF INVENTION: SITE, A METHOD FOR IMPROVING PROTEIN FUNCTION, AND A
 ; TITLE OF INVENTION: FUNCTION-MODIFIED PROTEIN
 ; FILE REFERENCE: 2003-0049
 ; CURRENT APPLICATION NUMBER: US/10/345,205
 ; CURRENT FILING DATE: 2003-03-16
 ; PRIOR APPLICATION NUMBER: 09/697,138
 ; PRIOR FILING DATE: 2000-10-27
 ; PRIOR APPLICATION NUMBER: 09/355,486
 ; PRIOR FILING DATE: 1999-09-20
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 6
 ; LENGTH: 775
 ; TYPE: PRT
 ; ORGANISM: Unknown
 ; FEATURE:
 ; OTHER INFORMATION: Description of Unknown Organism: DNA polymerase
 US-10-345-205-6

Query Match 29.8%; Score 1246.5; DB 15; Length 775;
 Best Local Similarity 36.5%; Pred. No. 3.8e-106;
 Matches 304; Conservative 150; Mismatches 282; Indels 97; Gaps 23;
 QY 8 VLDSSYEVVKGPEOVIIINGAENGVRVULIDRFRFVYFVALLAPAGDPKQV---AORIR 63
 Db 2 VLDVDYITEBCKPVIRLPK-KENGKPKIEHDFTRFYIYALLDSDSKIEVKKIGERHG 60
 QY 64 ALSRPKSPITGVDEKCKYFGRPRRURIKTVLPEAVREYRELVNRNVDGVLEADIRP 123
 Db 61 KIVR----IVDVEKVERKFLKPIITWKLYLEHPQDVPFTRKXVREHPHVAVDIFEDIP 116
 QY 124 AMRYLIDHDLPFTWYRVEAEPLKNGKGFVVDVILVKSRPPLVGEALAPTKLPDLRL 183
 Db 117 AKRYLIDKGLI-----PMGE-----BELKIL 138
 QY 184 AFDIEVSKQSPRPDPVIVIAVKTDDGDEVL-----FIAEGKDRKPIREFVEY 235
 Db 139 AFDIEVHEGE-EFAGKPIIMISYADSEAKVITWKIDLPFYEVVSSEREMIKRFLKV 197
 QY 236 VRRYDPDIIIVGNNHDFWFLRRARILGILKLDVTRVGAEPPTTSVHGH---VSVPGRL 292
 Db 198 IREKDPDVIITNGSDFDLPYLKGAELGILKPLGRD-GSEPKQRLGDMTAVEIKGRI 256
 QY 293 NVLDYDABEMPEIKISLEVAEYLGVMKSERVINWMEIPDYWDPPKRLDLOQYAR 352
 Db 257 HFDLYHVRITINLPTYLEAVYSAI-FGKPKKYYAH--EIAEAMTGTGLERVAKYSM 313
 QY 353 DVVRATYGLAEKILPFAIQLSVVTGLPLDQVGMVGFRLWYLIRAAFMKELVNRVE 412
 Db 314 EDAKVITYELGREFPMEAGLSLNGQPLMDVSRSTGNLVEMYLKAYERNELAKPKD 373
 QY 413 RPE-----ETVRGAILVLEPLRGVHENTAVLDPSMYPNIMIKNVGPDILVRPGEKCGE 466
 Db 374 EREYERRLRESYAGAYVKEPEKGLWGLVSLDPRSLYPSIIITHNVSPDTLNRE-----428
 QY 467 CGCWE---APEVKHFRFCPPGFFKTVLERLLELRKSVRAEMKKYPPDSPEYRLDDEKQ 525
 Db 429 CKNDIAPVGHKFCQDIPGFIPLSLKRLDDEKQEKRAKWA-TQDPIELKLDYRQAI 487
 QY 526 KVLNANSYGYVMGSGARWYCRCAKAVTANGRHILIR-TAINIARKLGLKVYGDPSL 584
 Db 488 KLLANSFYGYGYAKARWYCKEASVTANGREYIEFVRKELEERFGFKVLYIDTDL 547
 QY 585 TYD-----PEKVENFIKIKKEELG--FEIKLEKYKRLFFTEAKKRYAGLEDPRI 635
 Db 548 TIPGSESEETKKALEFVKYINSKLPGLLELEYEGFYKRGFFV-TKKRYAVIDEKGVIT 606
 QY 636 VGFVAVRGDMCELAKVQTKVBEIVLKTSEVNKAVEYVAKIVKELEBEGKVPLEKLV 695
 Db 607 RGLIEVRDWESEIAKSTQAKVAILKHGNVBEAVKIVKEVTEKLSKYEIPPEKLV 666
 QY 696 LSKLLEBYTTEAHPHVAAKMELSGVYRSPGDKIGVIVKGGRIISQRAWPFYVVKD 752
 Db 667 ITRPLHEYKALGHVAVAKKLAARGVKVPGMVIGIIVLRGDPISKRA---ILAEF 722
 QY 753 PS---QIDVTYVVDHQIIPAAIRILGYFGITEKKLKAASATQKTLDFLAKSK 802
 Db 723 PKCHKYDAEYIENQVLPVAVLRILEGFGYRKEKEDLRWQKTKQTGLTAWLNKKK 775

RESULT 15

US-10-227-110-29
 ; Sequence 29, Application US/10227110
 ; Publication No. US20030143577A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Stratagene
 ; TITLE OF INVENTION: HIGH FIDELITY DNA POLYMERASE COMPOSITIONS AND USES THEREFOR
 ; FILE REFERENCE: 25436/2155C
 ; CURRENT APPLICATION NUMBER: US/10/227,110
 ; CURRENT FILING DATE: 2002-08-23
 ; PRIOR APPLICATION NUMBER: 10/079,241

